

SecGraph: Towards SGX-based Efficient and Confidentiality-Preserving Graph Search (full version)*

Qiu hao Wang, Xu Yang, Saiyu Qi^(✉), and Yong Qi

Xi'an Jiaotong University, Xi'an, Shaanxi, China
 {QiuHaoWang,yangxu}@stu,xjtu.edu.cn, {saiyu-qi,qiy}@xjtu.edu.cn

Abstract. Graphs have more expressive power and are widely researched in various search demand scenarios, compared with traditional relational and XML models. Today, many graph search services have been deployed on a third-party server, which can alleviate users from the burdens of maintaining large-scale graphs and huge computation costs. Nevertheless, outsourcing graph search services to the third-party server may invade users' privacy. PeGraph was recently proposed to achieve the encrypted search over the social graph. The main idea of PeGraph is to maintain two data structures *XSet* and *TSet* motivated by the OXT technology to support encrypted conductive search. However, PeGraph still has some limitations. First, PeGraph suffers from high communication and computation costs in search operations. Second, PeGraph cannot support encrypted search over dynamic graphs. In this paper, we propose an SGX-based efficient and confidentiality-preserving graph search scheme *SecGraph* that can support insertion and deletion operations. We first design a new proxy-token generation method to reduce the communication cost. Then, we design an LDCF-encoded *XSet* based on the Logarithmic Dynamic Cuckoo Filter to reduce the computation cost. Finally, we design a new dynamic version of *TSet* named *Twin-TSet* to enable encrypted search over dynamic graphs. We have demonstrated the confidentiality preservation property of *SecGraph* through rigorous security analysis. Experiment results show that *SecGraph* yields up to $208\times$ improvement in search time compared with PeGraph and the communication cost in PeGraph is up to $540\times$ larger than that in *SecGraph*.

1 Introduction

Graphs are gaining increasing attention since they have expressive power and play an important role in many applications, such as social networks [15], biological data analyses [19], recommender systems [10], etc. Essentially, this is because the core data involved in these applications can be conveniently represented as graphs. For example, social networks (e.g., Facebook [1] and Instagram [2]) constitute various social users, essentially a graph where vertices represent users and

* This paper has been accepted by DASFAA 2024.
 The first two authors contributed equally and share the “co-first author” status.

edges represent their relationships, such as friendship. The wide use of graphs has brought about the emergence of graph search, such as shortest path search [11] and neighbor query [18], etc. As the scale of the graph surges, graph owners (e.g., enterprises for graph based services) desire to outsource their graphs to a third-party server, which however may snoop sensitive user information, e.g., users’ social connections, interests, and potentially sensitive information.

To enable encrypted graph search, some solutions rely on structure encryption [3] to enable private adjacent vertices search over encrypted graph [5, 12, 14]. However, these solutions just support single keyword search (i.e., search for the neighbor vertices of a vertex). Recently, Wang *et al.* [16] proposed PeGraph based on the OXT technology [4] to support encrypted conductive search over social graphs, i.e., search for common neighboring vertices (e.g., friends) of multiple vertices (e.g., people). Specifically, PeGraph maintains a multimap data structure $TSet$ to store the encrypted form of mapping from vertices to all their neighbor vertices and a set $XSet$ to store pairs of adjacent vertices information at the server side. Upon receiving search keywords (e.g., $w_1 \wedge \dots \wedge w_n$), the server first finds the matching neighbor vertices of w_1 in $TSet$ as an initial search result. Then, for each of its neighbor vertex, the server checks the existence of all pairs where each one is made up by it with a remaining search keyword $w_i, i \in [2, n]$, in $XSet$ to determine whether to insert it to the final result.

However, PeGraph suffers from three limitations. First, PeGraph incurs high communication costs in a search operation since it requires two search roundtrips between a client and the server, due to the characteristic of the OXT technology [4]. Second, PeGraph incurs high computation cost since it requires $c \cdot (n-1)$ expensive exponentiation modulo operations to check the existence of each neighbor vertex in the initial search result, where c is the number of neighbor vertices in the initial search result and n denotes the search keywords count. Last, PeGraph cannot support encrypted search over dynamic graphs since the OXT technology is designed to be applied to static encrypted search.

In this paper, we propose an *SGX-based efficient and confidentiality-preserving graph search scheme (SecGraph)* to provide secure and efficient search services over encrypted dynamic graphs. To address the first limitation, we design a *proxy-token generation method* by exploiting the trusted hardware SGX as the client’s trusted proxy on the server side to reduce the communication cost of search operations. To address the second limitation, we design a novel data structure *LDCF-encoded XSet* based on the Logarithmic Dynamic Cuckoo Filter [17] to transform the expensive exponentiation modulo operations required for existence checking operations in the OXT technology into the LDCF-based membership check process which can be conducted within storage-constraint SGX to reduce the computation cost. To address the last limitation, we design a new dynamic version of $TSet$ named *Twin-TSet* to record the relationships between adjacent vertices to efficiently handle encrypted searches over dynamic graphs. In summary, the contributions are the following.

- We propose an SGX-based efficient and confidentiality-preserving graph search scheme *SecGraph* to provide secure and efficient graph search services.

In this paper, we assume that the search keywords count issued by a client is n in the form of $w_1 \wedge \dots \wedge w_n$ and w_1 is the least frequent unless otherwise specified.

- In *SecGraph*, we design a new proxy-token generation method to reduce the communication cost. Then, we design an LDCF-encoded *XSet* to reduce the computation cost. Moreover, we design a new dynamic version of *TSet* named *Twin-TSet* to enable encrypted search over dynamic graphs.
- We further extend *SecGraph* into two optimized schemes *SecGraph-G* and *SecGraph-P* by adopting the fingerprint grouping and checking parallelization strategies, respectively, to speed up the search procedure.
- Finally, experiment results show that *SecGraph*, *SecGraph-G* and *SecGraph-P* respectively yield up to $208\times$, $572\times$ and $3,331\times$ improvements in search time compared with the state-of-the-art scheme PeGraph, and the communication cost in PeGraph is up to $540\times$ larger than that in *SecGraph*.

2 Related Work

Graphs are widely used to model structure data in various domains (e.g., social networks [15], and protein structures [19], and more) due to their excellent capability in characterizing the complex interconnections among entities. Graph search, as one of the fundamental tasks in graph analytics, has gained increasing attention in recent years. Numerous algorithms have been proposed to address diverse types of graph search, such as graph similarity search [6] and shortest path search [11], etc. However, all of them focus on graph search in the plaintext domain without considering privacy preservation.

Along with the development of the data outsourcing paradigm, enabling efficient search over encrypted outsourced graphs has attracted widespread attention from both academia and industry. Structure encryption [3] is a promising cryptographic technique to provide private adjacent vertices search over encrypted graphs. For example, Chase *et al.* [5] firstly extended the notion of structured encryption to the setting of arbitrarily structured data including complex graph data. Based on this, Shen *et al.* [14] proposed a privacy-preserving scheme PP k NK to achieve top- k nearest keyword search on graphs. Furthermore, Lai *et al.* [12] proposed a privacy-preserving graph scheme GraphSE² to facilitate parallel and encrypted graph data access on large social graphs. Recently, Wang *et al.* [16] proposed PeGraph using the OXT technique [4] and a more lightweight secure multi-party computation method to expand the richness of search and improve performance. PeGraph is state-of-the-art privacy-preserving social graph search scheme. However, there are some drawbacks. First, PeGraph incurs high communication costs since it requires two search roundtrips between a client and the server during the search procedure. Second, PeGraph suffers from high computation costs since it requires expensive exponentiation modulo operations to check the existence of each entry in the initial search result. Last, PeGraph cannot support encrypted search over dynamic graphs since the OXT technology is applied to static encrypted search. In this paper, we devote ourselves to achieve an efficient and confidentiality-preserving graph search scheme while supporting dynamic graph updates.

3 Preliminaries

Intel SGX and Enclave. Intel SGX [7] is a set of extensions of x86 instructions that provides trusted execution environments (i.e., *enclave*) to protect the

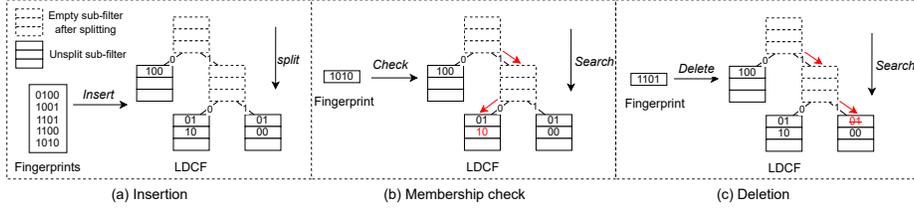


Fig. 1. Illustration of the data structure LDCF and three associated operations.

integrity and confidentiality of the application data and the code. The enclave is limited to 128MB to store data. If this limit is exceeded, the enclave will automatically apply the page-swapping mechanism, causing severe performance degradation. The enclave has three main security properties: (1) *isolation*: any software outside the enclave can not directly access code or data within it, but it can access the entire virtual memory of its untrusted host; (2) *sealing*: it enables the process of encrypting and authenticating the data within it; (3) *attestation*: there is a secure channel between an external party and the enclave.

Oblivious Cross-Tags Protocol. Oblivious cross-tags protocol (OXT) [4] is a searchable encryption technique designed for secure and efficient conjunctive search over text files. Generally, the core idea of OXT is maintaining a multimap data structure $TSet$ and a set data structure $XSet$ at the server side, $TSet$ records a set of pairs (fid, y) for each keyword w labeled by a corresponding value $stag(w)$ where fid is the encrypted file identifier and y is a blinded value, $XSet$ records a list of values $xtag(w, fid)$ over each keyword-file-identifier pair. Specifically, given a pseudo-random function (PRF) F and the keys (K_I, K_X, K_T, K_Z) , we assume there are c files $\{fid_i\}, i \in [1, c]$ containing the keyword w . The client sequentially calculates $stag(w) = F(K_T, w)$, $y = F(K_I, fid_i) \cdot F(K_Z, w || i)^{-1}$, and $xtag(w, fid_i) = g^{F(K_X, w) \cdot F(K_I, fid_i)}$, $i \in [1, c]$, where g is the generator of a cyclic group and $||$ represents the concatenation operation. When issuing a conjunctive search (e.g., $w_1 \wedge \dots \wedge w_n$), the client first sends the search token $stag(w_1)$ to retrieve all matching encrypted fid containing w_1 from the server. With the size c of the initial search result just retrieved, the client then generates and sends the intersection tokens $xtoken(w_1, w_i, j) = g^{F(K_Z, w_1 || j) \cdot F(K_X, w_i)}$, $i \in [2, n], j \in [1, c]$ to the server. Upon receiving them, for each encrypted fid in the initial search result, the server will determine whether to insert it into the final search result by checking $xtoken(w_1, w_i, j)^y = g^{F(K_X, w_i) \cdot F(K_I, fid_j)} \stackrel{?}{=} xtag(w_i, fid_j)$, $i \in [2, n]$ for each $j \in [1, c]$. Using the OXT technology, the server can exactly find the search result without knowing either file identifier fid or keywords $w_i, i \in [2, n]$.

Cuckoo Filter and Dynamic Cuckoo Filter. Cuckoo Filter (CF) [9] is a compact data structure that enables approximate set membership checks in static settings. The main idea of the CF is each element is hashed to one or more buckets, and these buckets store fingerprint information derived from the hashed values, where the fingerprint is a small portion of the original hash value, typically a few bits. To support efficient inserting, deleting, and checking an item in the CF, Zhang *et al.* [17] extend the traditional CF to a novel data structure Logarithmic Dynamic Cuckoo Filter (LDCF) where a fully inserted CF is divided into two CFs in a binary tree shape recursively based on the prefix of the

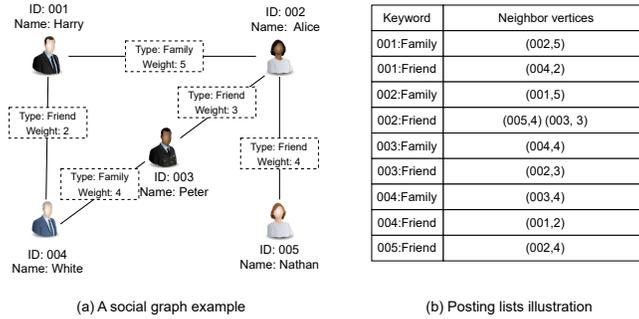


Fig. 2. A toy example for illustrating a social graph.

fingerprints, as shown in Fig. 1. Due to the features of the tree structure, LDCF achieves a sub-linear search and reduces space storage overhead by omitting the storage space of common prefixes for the fingerprints during the process of tree expansion. For convenience, we refer to each CF in LDCF as a *sub-filter* in this paper. LDCF consists of three algorithms:

- **LDCF.Insert(x):** Upon receiving an inserting element x , the algorithm first computes its fingerprint δ and uses it to locate the matching *sub-filter*. Then it calculates two candidate positions μ and ν and finally puts δ in one of the two candidate positions in the matching *sub-filter*. Fig. 1 (a) shows the procedure of inserting five fingerprints into an LDCF.
- **LDCF.Membershipcheck(x):** Upon receiving a query element x , the algorithm first calculates its fingerprint δ and finds the matching *sub-filter*. Then, it derives the corresponding positions μ and ν to check whether δ exists in one of them and returns true if exists, otherwise false. Fig. 1 (b) shows the procedure of checking a fingerprint whether exists in the LDCF.
- **LDCF.Delete(x):** Upon receiving a deleting element x , the algorithm first calculates its fingerprint δ and finds the matching *sub-filter*. Then, it derives the corresponding positions μ and ν to locate δ , and finally deletes it and returns true, otherwise false. Fig. 1 (c) shows the procedure of deleting an existing fingerprint in the LDCF.

4 System Overview

4.1 Problem Definition

Formally, the social graph can be represented by $G = (V, E)$, where V is a vertex set in the graph and E is a relationship set among vertices. As shown in Fig. 2(a), each $v \in V$ represents a vertex identified by id , it also has an attribute $name$, e.g., the vertex '001' named 'Harry'. Each $e \in E$ represents an edge between vertices identified by $type$, with $weight$ representing the importance of the edge, e.g., there is a friendship relationship between vertices '001' and '002', with an importance of 5. A relationship can be defined by $(id_{out}, id_{in}, type)$, which means the edge labeled by $type$ starts from id_{out} and ends with id_{in} . As shown in Fig. 2(b), our system maintains a posting list for each vertex. The posting list is a map data structure, where the keyword is $id_{out} : type$ and the

value is composed of a set of $(id_{in}, weight)$ pairs, where id_{in} is the identifier of the neighbor vertex of id_{out} , and $weight$ represents the importance of the relationship, e.g., the intimacy between friends. In this paper, we denote id_{out} as w , and the search types on graphs are divided into two types: (1) *Exact search*. Search for neighbor vertices with a specific relationship to multiple vertices, e.g., the common friend of vertices '003' and '005' is vertex '002'. (2) *Fuzzy search*. Search for vertices whose *name* contain the same sub-string, e.g., the vertices with 'ha' in their *name* include vertices '001' and '005'.

4.2 System Model

There are two entities in the system: the client and the SGX-enabled server. (1) *Client*. The client launches a remote attestation and establishes a secure channel with the trusted part (i.e., enclave) of the server, then sends a set of secret keys to the enclave. The client can securely insert or delete edges or vertices, and issue conjunctive search. (2) *SGX-enabled server*: The server will insert or delete the corresponding encrypted index and provide graph search services for the client. We assume that the client is absolutely honest but the server is honest-but-curious. That is, the server will execute the established program correctly, but powerful enough to get full access over the software stack (such as OS and hypervisor) outside of the enclave, and can infer sensitive information from encrypted data by observing search tokens and search results. We state that a series of side-channel attacks against SGX are out of our scope.

In this paper, *SecGraph* aims to achieve three goals in terms of efficiency, functionality, and security : (1) *Efficient search*. *SecGraph* should provide efficient search while reducing communication and computation costs. (2) *Provision dynamic update*. *SecGraph* should support encrypted search over dynamic graphs. (3) *Confidentiality preservation*. *SecGraph* should protect the confidentiality of search keywords and results.

5 Detailed Construction of *SecGraph*

5.1 Design Rationale of *SecGraph*

As mentioned, Wang *et al.* [16] proposed PeGraph that can support versatile search over encrypted social graphs. The main idea of PeGraph is combining the OXT technology [4] and the additive secret sharing [8] to provide private ranked conjunctive search. Specifically, PeGraph first extracts each keyword w in the form of $id_{out} : type$ from graphs and associates the posting list $\{(id_{in}, weight)\}$ with w . Then, for each w , PeGraph calculates $stag(w)$ to index its corresponding posting list. In addition, for each id_{in} in the set $\{(id_{in}, weight)\}$, PeGraph calculates $xtag(w, id_{in})$ and adds it to $XSet$, which allows the server to check the existence of an id_{in} without knowing either w or id_{in} . After obtaining the matching id_{in} , PeGraph further allows the server to obliviously render the encrypted search results in a ranked order according to their importance $weight$ by using the additive secret sharing primitive and returning the top- k results, avoiding unnecessary downlink traffic.

Despite its considerable performance, PeGraph still has some problems due to the characteristics of OXT technology [4]. First, PeGraph incurs high communication costs since it requires two search roundtrips between a client and the server during the search procedure. The reason is that the OXT technology requires the client to retrieve the encrypted matching posting list associated with a certain keyword (e.g., w_1) as an initial search result from the server in advance to assist in generating intersection tokens (e.g., $xtoken(w_1, w_i, j), i \in [2, n], j \in [1, c]$). Therefore, the first challenge is how to achieve one search roundtrip to reduce the communication cost. Second, PeGraph suffers from high computation cost since it requires $c \cdot (n-1)$ expensive exponentiation modulo operations to check the existence of each id_{in} in the initial search result, where c is the number of id_{in} in the initial search result and n is the number of search keywords. Therefore, the second challenge is how to check the existence of id_{in} without using expensive exponentiation modulo operations to reduce computation cost. Finally, PeGraph cannot support encrypted search over dynamic graphs since the OXT technology only supports static encrypted search. Accordingly, the third challenge is how to enable dynamic encrypted graph search.

To address the first challenge, we propose a *proxy-token generation method* to avoid the procedure of returning the initial search result to the client. The main idea is to resort to SGX as the client’s trusted proxy at the server side to generate a search token $stag(w_1)$ and intersection tokens $xtoken(w_1, w_i, j), i \in [2, n], j \in [1, c]$ due to the trusted computing power of SGX. Specifically, when issuing a graph search, a client sends the search keywords $w_1 \wedge \dots \wedge w_n$ to the enclave (i.e., the trusted part of the server) via a secure channel, and the latter generates $stag(w_1)$ and retrieves the initial search result from the untrusted part of the server to generate intersection tokens independently. In this way, only one search roundtrip is required between the client and the server, which largely reduces the communication cost.

To address the second challenge, we start with an observation from the OXT technology [4]. The server needs to perform a large number of expensive exponentiation modulo operations to check whether each $xtag(w_i, id_{in}), i \in [2, n]$ is in $XSet$ or not to determine whether id_{in} belongs to the final search result, which can incur significant performance degradation. Fortunately, we note that the procedure of checking the existence of each id_{in} in the initial search result is essentially a membership check process. Thus, a naive solution is to load the data structure $XSet$ into the enclave due to the trusted storage power of SGX and allow the enclave to check whether each $xtag(w_i, id_{in}), i \in [2, n]$ exists in $XSet$ in plaintext rather than perform exponentiation modulo operation over ciphertexts. However, as the entire keywords in graphs are encoded into $XSet$, the size of $XSet$ will become large and may not be stored in the enclave due to its limited storage, which will incur the page-swapping mechanism and also cause performance degradation. To do this, we use a compact data structure Logarithmic Dynamic Cuckoo Filter (LDCF) [17] to store $XSet$. The *LDCF-encoded $XSet$* not only allows the enclave to load a few *sub-filters* to complete the procedure of checking the existence of each id_{in} in the initial search result instead of loading the entire *sub-filter* set but provides a sub-linear search to locate the matching *sub-filters* quickly. In this way, using the LDCF-encoded $XSet$, the computation cost is largely reduced during the search procedure.

To address the last challenge, we note that applying the OXT technology [4] to the dynamic setting is not straightforward since the pre-computed multimap data structure $TSet$ does not need to change with time, thus it is impossible to emulate in the dynamic setting, where the graphs are continuously updated (including insertion and deletion). To do this, we design a new dynamic version of $TSet$ named $Twin-TSet$ that contains a pair of map data structures, one is $TSet$ storing the relationship between a pair made up by a keyword w and an associated update count c and a neighbor vertex id_{in} the other is $ITSet$ (i.e., inverse $TSet$) storing the relationship between a pair made up by a keyword w and a neighbor vertex id_{in} and the corresponding update count c . When inserting a pair (w, id_{in}) , we first increase the update count c associated with w by one and then insert the two pairs $((w, c), id_{in})$ and $((w, id_{in}), c)$ into $TSet$ and $ITSet$ in the form of ciphertext respectively. When deleting a pair (w, id_{in}) , we first retrieve the corresponding update count c from $ITSet$ and then replace the id_{in} at position (w, c) of $Tset$ with the latest id'_{in} at position (w, c') , where c' is the latest update count associated with w . After that, we accordingly insert a new pair $((w, id'_{in}), c)$ into $ITSet$ and decrease c by one. when searching w , similarly with the OXT technology, we only need to retrieve c values from $TSet$ at positions $(w, i), i \in [1, c]$ without considering deleted identifiers id_{in} since they have been deleted previously. At present, it seems that the $Twin-TSet$ can support encrypted search on dynamic graphs, but how to store the update count for each w privately is still an issue. Fortunately, we can protect it from being leaked to the server by storing a map data structure $UpdateCnt$ within the enclave with its trusted storage power.

Until now, we have illustrated the design rationale of the proposed SGX-based efficient and confidentiality-preserving graph search ($SecGraph$). It is worth noting that, due to the trusted computing power of SGX, $SecGraph$ can also support returning top- k ranked search results to the client by directly decrypting and sorting the neighbor vertice identifiers according to their *weight* without performing additive secret sharing operations over the encrypted posting lists.

5.2 Design Details of $SecGraph$

Now we proceed to describe our construction of $SecGraph$ in detail. $SecGraph$ utilizes a set of PRFs $(F_1, F_2, F_3 : \{0, 1\}^\lambda \times \{0, 1\}^* \rightarrow \{0, 1\}^\lambda)$, two hash functions $H_1 : \{0, 1\}^* \rightarrow \{0, 1\}^\alpha$ and $H_2 : \{0, 1\}^* \rightarrow \{0, 1\}^\xi$. The workflow of $SecGraph$ can be divided into the following protocols.

SetUp Protocol. This protocol is responsible for initializing some secret keys and data structures. Specifically, the client launches a remote attestation and establishes a secure channel with the enclave first. Then it generates secret keys (K_T, K_Z, K_X) for PRFs (F_1, F_2, F_3) and sends them to the enclave. The enclave initializes three empty data structures: (1) $UpdateCnt$ that stores the number of id_{in} corresponding to each w ; (2) CFs that stores a mapping from $sub-filtersId$ to cache $sub-filters$; (3) $IndexTree$ that stores the split state of LDCF-encoded $XSet$ to locate the matching $sub-filters$ during the update and search procedures. The server initializes three empty data structures: (1) $TSet$ that stores the encrypted posting list; (2) $ITSet$ that stores a mapping from the neighbor vertices' id to their location in the posting list to locate the deleting neighbor vertices during the deletion procedure; (3) $XSet$ that stores the fingerprints of all $xtag$.

Algorithm 1: Update

```

Input: Secy keys  $K_T, K_Z, K_X$ , update counter  $UpdateCnt$ , updated triplet  $(w, id, weight)$ , update operation  $op$ , and  $IndexTree$ .
Output: Encrypted database  $EDB=(TSet, ITSet, XSet)$ .
Client:
1: Send  $(w, id, weight), op$  to the enclave;
Enclave:
2:  $xtag = (w||id), \delta = H_2(xtag), \mu = H_1(xtag)$ ;
3: Get  $sub-filterId$  according to  $\delta$  and  $IndexTree$ ;
4: if  $op = insert$  then
5:   if  $UpdateCnt[w] = \perp$  then
6:      $UpdateCnt[w] = 0$ ;
7:      $UpdateCnt[w]++$ ;
8:      $stag = F_1(K_T, w||UpdateCnt[w])$ ;
9:      $C_{id} = (id||weight) \oplus F_2(K_Z, w)$ ;
10:     $ind = F_3(K_X, w||id)$ ;
11:     $C_{stag} = (w||UpdateCnt[w]) \oplus F_2(K_Z, w)$ ;
12:    Send  $(sub-filterId, stag, C_{id}, ind, C_{stag}, \delta, \mu, op)$  to the server;
13:  else
14:     $stag^{-1} = F_1(K_T, w||UpdateCnt[w])$ ;
15:     $UpdateCnt[w]--$ ;
16:     $C_{id}^{-1} = TSet[stag^{-1}]$ ; /* ocall */
17:     $(id^{-1}||sort_i^{-1}) = C_{id}^{-1} \oplus F_2(K_Z, w)$ ;
18:     $ind^{-1} = F_3(K_X, w||id^{-1}), ind = F_3(K_X, w||id)$ ;
19:     $C_{stag} = ITSet[ind]$ ; /* ocall */
20:     $(w||c) = C_{stag} \oplus F_2(K_Z, w)$ ;
21:     $stag = F_1(K_T, w||c)$ ;
22:    Send  $(sub-filterId, stag, stag^{-1}, ind, ind^{-1}, \delta, \mu, op)$  to the server;
Server:
23: if  $op = insert$  then
24:    $TSet[stag] = C_{id}, ITSet[ind] = C_{stag}$ ;
25:    $XSet[sub-filterId].Insert(\delta, \mu)$ ;
26: else
27:    $TSet[stag] = TSet[stag^{-1}]$ ;
28:    $ITSet[ind^{-1}] = ITSet[ind]$ ;
29:   Delete  $ITSet[ind]$ ;
30:    $XSet[sub-filterId].Delete(\delta, \mu)$ ;
31: Update  $IndexTree$  if necessary; /* ecall */

```

Update Protocol. This protocol allows the client to issue graph updates. At the beginning, the client generates an update request $(w, id, weight, op)$, where op is the operation type (insertion or deletion) (line 1). Upon receiving it, the enclave first calculates $xtag$ using w and id and generates its fingerprint δ and candidate index μ (line 2). Then, the enclave searches the path of the matching $sub-filter$ $sub-filterId$ in $IndexTree$ based on δ (line 3). Specifically, the enclave traverses each bit of δ to deeply search the $IndexTree$ by selecting the left subtree when the bit is 0, otherwise selecting the right subtree, until finding a leaf node, and the traversed δ sub-string is the path of the $sub-filter$ $sub-filterId$. If $op = insert$, the enclave increases the $UpdateCnt[w]$ by 1, then encrypts the $(w||c, id)$, $(w||id, c)$ pairs in forms of $(stag, C_{id})$ and (ind, C_{stag}) for $TSet$ and $ITSet$ respectively. Finally, the enclave sends $(\delta, \mu, sub-filterId)$ and the pairs to the server (lines 4-12). If $op = delete$, the enclave first calculates w 's latest id index $stag^{-1} = F_1(K_T, w||UpdateCnt[w])$ and decreases $UpdateCnt[w]$ by 1, then loads the latest encrypted id (i.e., C_{id}^{-1}) and encrypts it (lines 14-17). Next, the enclave computes the indexes of the latest id (i.e., ind^{-1}) and the deleting id (i.e., ind) in $ITSet$ (line 18). Then, the enclave gets the deleting id 's encrypted index in $TSet$ (i.e., C_{stag}) from $ITSet[ind]$ and decrypts it to $stag$ (lines 19-21). Finally, the enclave sends $(stag, stag^{-1}, ind, ind^{-1}, \delta, \mu, sub-filterId)$ to the server (line 22). Upon receiving it from the enclave, if $op = insert$, the server will insert the $(stag, C_{id})$ and (ind, C_{stag}) pairs into $TSet$ and $ITSet$ respectively and insert the fingerprint δ into the $XSet$ (lines 23-25). If $op = delete$, the server will overwrite the deleting id with the latest id in $TSet$ and $ITSet$ then delete the fingerprint δ and deleting ind in $XSet$ and $ITSet$ respectively to remove the corresponding id (lines 26-30). Finally, if the LDCF-encoded $XSet$ is split, the server will execute an *ecall* operation to update the $IndexTree$ (line 31).

Search Protocol. This protocol allows the client to issue conjunctive graph searches. Specifically, the client selects search keywords $(w_1 \wedge \dots \wedge w_n)$ and sends them to the enclave (line 1). Upon receiving them, the enclave first initializes two empty lists $stokenList$ and $CResList$ to store w_1 's $stag$ and final encrypted results respectively (line 2). Then, the enclave traverses $UpdateCnt[w_1]$ to com-

Algorithm 2: Search

```

Input: Secrity keys  $K_T, K_Z, K_X$ , update
counter  $UpdateCnt$ , search token
 $(w_1 \wedge \dots \wedge w_n)$ , encrypted database
 $EDB=(TSet, ISet, XSet)$ .
Output: search result  $ResList$ .
Client:
1: Send  $(w_1 \wedge \dots \wedge w_n)$  to the enclave;
Enclave:
2: Initialize two empty lists  $tokenList$ ,
 $CResList$ ;
3: for  $j = 1$  to  $UpdateCnt[w_1]$  do
4:    $stag = F_1(K_T, w_1||j)$ ;
5:   Insert  $stag$  into the  $tokenList$ ;
6: Send  $tokenList$  to the server;
Server:
7: Initialize an empty list  $CidList$ ;
8: for  $j = 1$  to  $tokenList.size$  do
9:    $Cid_j = TSet[tokenList[j]]$ ;
10:  Insert  $Cid_j$  into the  $CidList$ ;
11: Send  $CidList$  to the enclave;
Enclave:
12: Initialize an empty map  $CFs$ ;
13: for each  $Cid_j \in CidList$  do
14:    $flag = 0$ ;
15:    $(id_j||weight_j) = Cid_j \oplus F_2(K_Z, w_1)$ ;
16:   for  $i = 2$  to  $n$  do
17:      $xtag_{i,j} = (w_i||id_j), \delta = H_2(xtag_{i,j})$ ;
18:     Get  $sub-filterId$  according to  $\delta$  and
 $IndexTree$ ;
19:     if  $sub-filterId$  not in  $CFs$  then
20:       Load matching  $sub-filter$ ; /* ocall */
21:        $CFs[sub-filterId] = sub-filter$ ;
22:        $\mu = H_1(xtag_{i,j}), \nu = \mu \oplus H_1(\delta)$ ;
23:       if  $\delta$  not in  $CFs[sub-filterId][\mu]$  and
 $CFs[sub-filterId][\nu]$ 
24:          $flag = 1, break$ ;
25:       if  $flag = 0$  then
26:         Insert  $Cid_j$  into the  $CResList$ ;
27: Send  $CResList$  to the client;
Client:
28: Initialize an empty list  $ResList$ ;
29: for each  $Cid_j \in ResList$  then
30:    $(id_j||weight_j) = Cid_j \oplus F_1(K_T, w_1)$ ;
31:   Insert  $id_j$  into the  $ResList$ ;

```

pute all of $stags$ for w_1 and sends them to the server (lines 3-6). The server initializes an empty list $CidList$, and then traverses the $tokenList$ to get the encrypted id (i.e., Cid) of w_1 from $TSet$ and returns them to the enclave (lines 7-11). For each Cid_j , the enclave decrypts it to get id_j and calculates the $xtag_{i,j}$ using $\{w_i\}, i \in [2, n]$ (lines 13-17). To hide the $xtag$'s fingerprint from the server, the enclave generates the fingerprint δ and finds the corresponding $sub-filterId$ from $IndexTree$. If the matching $sub-filter$ doesn't exist in the cache table CFs , the enclave will execute an *ocall* operation to load the matching $sub-filter$ (lines 18-21). After that, the enclave computes two candidate positions μ and ν to check whether the δ exists in the $sub-filter$. Only all $xtag_{i,j}, i \in [2, n]$ pass the membership check id_j can be added into the final result $CResList$ (lines 19-26). Finally, the enclave returns $CResList$ to the client (line 27), and the client can obtain all decrypted id_j in $ResList$ (lines 28-31).

5.3 Extension to Fuzzy Search

SecGraph can also support the fuzzy search (such as sub-string search), e.g., find users whose *name* contains 'Ha'. To enable this, instead of directly creating a *stag* for each user's *name*, our idea is to split the *name* into a set of pairs composed of a sub-string with fixed-length s and an integer. For example, assume the length of a sub-string is 2, 'Harry' can be split into $\{('H', 1), ('Ha', 2), ('ar', 3), ('rr', 4), ('ry', 5), ('y$', 6)\}$, where each integer part refers to the absolute position pos of the sub-string in the *name*, '#' is the start character and '\$' is the terminator. We only make a few changes to *SecGraph* to enable it to support fuzzy search. Specifically, in Update protocol (Alg.1), the client sends (w, id, pos, op) to the enclave, where w is a sub-string (line 1). Then the enclave calculates $xtag = (w||id||pos)$ (line 2) and $Cid = (id||pos) \oplus F_2(K_Z, w)$ (line 9). The posting list for each keyword w is a set of (id, pos) pairs, where id is the vertex whose *name* contains w and pos is the absolute position of w in the *name*. In Search protocol (Alg.2), the client sends the search keywords in the form of

$w_1 \wedge (w_2, \Delta_2) \wedge \dots \wedge (w_n, \Delta_n), i \in [2, n]$, where Δ_i is the relative position between w_i and w_1 (line 1). Upon receiving the *CidList* from the server, the enclave obtains $(id_j || pos_j) = C_{id_j} \oplus F_2(K_Z, w_1)$ (line 15). After that, for each (w_i, Δ_i) , the enclave calculates the $xtag_{i,j} = (w_i || id_j || pos_j + \Delta_i)$ (line 17).

6 Optimization

SecGraph-G: Fingerprint Grouping. We observe that there exists a performance bottleneck in *SecGraph* that much time may be spent on *ocall* operations when loading numerous *sub-filters* into the enclave during a search procedure. To solve the drawback, we propose an optimized scheme *SecGraph-G* to reduce the number of *sub-filters* to be loaded. The strategy is fingerprint grouping by generating the fingerprint of w and the fingerprint of $xtag(w, id_{in})$ and then joining them together to form the final fingerprint instead of just generating the fingerprint of $xtag(w, id_{in})$. In this way, the fingerprints generated by a vertex and each of its adjacent vertices are highly likely to be stored in the same *sub-filter* when constructing the LDCF-encoded *XSet*.

SecGraph-P: Checking Parallelization. Furthermore, we note that the current membership check process of *xtag* is serial which also seriously affects the search performance of *SecGraph* (lines 13-24 in Alg.2). With the observation that each membership check process is independent, we propose another optimized scheme *SecGraph-P* to speed up the membership check process during the search procedure. The strategy is checking parallelization by loading all matching *sub-filters* into the enclave first and then parallelizing to check whether each *xtag* exists or not. In this way, *SecGraph-P* only takes one membership check time to complete all required membership check processes, in the best-case scenario.

7 Security Analysis

Before presenting a formal security analysis to show the security guarantee of *SecGraph*, we first define the leakage functions and then use them to prove the security. In *Setup* protocol, *SecGraph* leaks nothing to the server except for the empty encrypted database *EDB*. Thus we have $\mathcal{L}^{Stp} = (|TSet|, |ITSet|, |XSet|)$, where $|TSet|$, $|ITSet|$ and $|XSet|$ are ciphertext lengths of data structures of *TSet*, *ITSet* and *XSet* respectively. In *Update* protocol, *SecGraph* leaks access on *TSet*, *ITSet* and *XSet*. Thus, we have $\mathcal{L}^{Updt} = (op, |TSet[stag]|, |ITSet[ind]|, |XSet[sub-filterId]|)$, where $op = insert/delete$ denotes the update operation, $TSet[stag]$ indicates the encrypted identifier to be inserted in *TSet* with its *stag*, $ITSet[ind]$ indicates the encrypted *stag* to be inserted in *ITSet* with its *ind* and $XSet[sub-filterId]$ indicates the fingerprint to be inserted in *XSet* with its *sub-filterId*. In *Search* protocol, *SecGraph* leaks the search token *tokenList* and access pattern on *TSet* when the server finds the matching entries in *TSet* associated with w_1 , defined as ap_{TSet} and on *XSet* when the enclave locates the desired *sub-filters*, defined as ap_{XSet} . Thus, we have $\mathcal{L}^{Srch} = (tokenList, ap_{TSet}, ap_{XSet})$. Following the security definition in [13], we give the formal security definitions.

Definition 1. Let $\Pi = (Setup, Update, Search)$ be our *SecGraph* scheme. Consider the probabilistic experiments $\mathbf{Real}_{\mathcal{A}}(\lambda)$ and $\mathbf{Ideal}_{\mathcal{A},S}(\lambda)$ with a probabilistic

polynomial-time (PPT) adversary and a stateful simulator that gets the leakage function \mathcal{L} , where λ is a security parameter. The leakage is parameterized by \mathcal{L}^{Stp} , \mathcal{L}^{Uptd} and \mathcal{L}^{Srch} depicting the information leaked to \mathcal{A} in each procedure.

Real $_{\mathcal{A}}(\lambda)$: The challenger initialises necessary data structures by running **Setup**. When inputting graphs chosen by \mathcal{A} , it makes a polynomial number of updates (i.e., addition and deletion). Accordingly, the challenger outputs the encrypted database $EDB=(TSet, ITSet, XSet)$ with **Update** to \mathcal{A} . Then, \mathcal{A} repeatedly performs graph searches. In response, the challenger runs **Search** to output the result to \mathcal{A} . Finally, \mathcal{A} outputs a bit.

Ideal $_{\mathcal{A},\mathcal{S}}(\lambda)$: Upon inputting graphs chosen by \mathcal{A} , \mathcal{S} initialises the data structures and creates encrypted database $EDB=(TSet, ITSet, XSet)$ based on \mathcal{L}^{Stp} , and passes them to \mathcal{A} . Then, \mathcal{A} repeatedly performs range queries. \mathcal{S} simulates the search results by using \mathcal{L}^{Uptd} and \mathcal{L}^{Srch} and returns them to \mathcal{A} . Finally, \mathcal{A} outputs a bit.

We say Π is \mathcal{L} -adaptively-secure if for any PPT adversary \mathcal{A} , there exists a simulator \mathcal{S} such that $|\Pr[\mathbf{Real}_{\mathcal{A}}(\lambda)=1] - \Pr[\mathbf{Ideal}_{\mathcal{A},\mathcal{S}}(\lambda)=1]| \leq \text{negl}(\lambda)$, where $\text{negl}(\lambda)$ denotes a negligible function in λ .

Theorem 1. (Confidentiality of *SecGraph*). Assuming (F_1, F_2, F_3) are secure PRFs and (H_1, H_2) are secure hash functions. *SecGraph* is \mathcal{L} -secure against an adaptive adversary and ensures forward security and Type-III backward security.

Proof. We model the PRFs and the hash functions as random oracles $\{\mathcal{O}_{F_1}, \mathcal{O}_{F_2}, \mathcal{O}_{F_3}, \mathcal{O}_{H_1}, \mathcal{O}_{H_2}\}$ and sketch the execution of the simulator \mathcal{S} . In **Setup** protocol, \mathcal{S} simulates the encrypted database based on \mathcal{L}^{Stp} , which has the same size as the real one. Specifically, it includes two dictionaries \mathcal{D}_1 and \mathcal{D}_2 and a set \mathcal{T} . \mathcal{S} further simulates the keys in the enclave by generating random strings $(\bar{k}_1, \bar{k}_2, \bar{k}_3)$, which are indistinguishable from the real ones. When the first graph search sample (w_1, w_2) is sent, \mathcal{S} generates simulated tokens $\tilde{t}_i = \mathcal{O}_{F_1}(\bar{k}_1 || w_1 || i)$ from c to 1 and c is the number of matched entries from \mathcal{L}^{Stp} . For each matching value α in \mathcal{D}_1 with the address \tilde{t}_i , another random oracles \mathcal{O}_{F_2} is operated as $\tilde{R} = \mathcal{O}_{F_2}(\bar{k}_2 || w_1) \oplus \alpha$ to obtain \tilde{R} inside, where \tilde{R} has the same length as the real one. With \bar{k}_3 , three random strings $\delta = \mathcal{O}_{H_2}(\bar{k}_3 || w_2 || \tilde{R})$, $\beta = \mathcal{O}_{H_1}(\delta)$ and $\gamma = \beta \oplus \sigma$ are calculated to check whether δ is in any one of the locations of the simulated set \mathcal{T} . If yes, \mathcal{S} adds \tilde{R} into the results. When a new triplet $(w, id, weight)$ is added, the results can also be simulated based on \mathcal{L}^{Uptd} . Due to the pseudorandomness of PRFs and the hash function, \mathcal{A} cannot distinguish between the tokens and results of **Real $_{\mathcal{A}}(\lambda)$** and **Ideal $_{\mathcal{A},\mathcal{S}}(\lambda)$** .

Following the definitions of forward and backward security in [13], we prove *SecGraph* achieves both forward security and Type-III backward security. Forward security is straightforward since the data structure *UpdateCnt* ensures that \mathcal{A} cannot generate search tokens to retrieve newly added identifiers when adding a new triplet. As for backward security, remembering when the entries in *ITSet* were added and deleted leaks when additions and deletions for w took place. Extracting all update counts and correlating them with the update timestamps

Forward security refers to newly inserted data is no longer linkable to searches issued before, and backward security refers to deleted data is no longer searchable in searches issued later [13].

Table 1. Summary of the graph data used in our experiments.

Dataset	Nodes	Edges	Edge type	Graph type	Source link
Email	36,692	183,831	Friendship	Undirect	snap.stanford.edu/data/email-Enron.html
Youtube	1,134,890	2,987,624	Exchange	Undirect	snap.stanford.edu/data/com-Youtube.html
Gplus	107,614	13,673,453	Share	Directed	snap.stanford.edu/data/ego-Gplus.html

reveals the specific addition that each deletion canceled. Nevertheless, the identifiers are encrypted by XORing a PRF value, the server cannot learn which identifiers contained w but have not been removed. Based on these leakages, *SecGraph* guarantees Type-III backward security. \square

8 Experimental Evaluation

8.1 Experiment Settings

In the experiments, we implement the state-of-the-art encrypted graph search scheme PeGraph [16], and our four schemes *SecGraph*, *SecGraph-G*, *SecGraph-P* and *SecGraph-A* (adopting both the fingerprint grouping and checking parallelization strategies) in about 5k LOCs of C++. The client and server are deployed on a workstation equipped with an SGX-enabled Intel(R) Core(TM) i7-10700 CPU@2.60GHz with Ubuntu 18.04 server and 64GB RAM. For cryptographic primitives, we use the cryptography library Intel SGX SSL and OpenSSL (v1.1.1n) to implement the pseudorandom function via HMAC-256 and use SHA-256 to generate hash values for fingerprint. For implementing the LDCF-encoded *XSet*, we adopt the open-source code of the LDCF provided in [17]. We use three real-world datasets Email, Youtube, and Gplus in our experiments, as shown in Table.1. All experiments were repeated 20 times and the average is reported.

8.2 Performance Evaluation

In our experiments, we default to setting the fingerprint length, the *sub-filter* size, and bucket size in *SecGraph* as 16 bits, 10,000, and 4, respectively, according to the recommendation of LDCF [17] since there is an acceptable balance between the accuracy and speed of membership check under these parameters.

Update Performance. We first evaluate the insertion performance of PeGraph and *SecGraph* as the used dataset size ratio increases from 20% to 100%. As we can see from Fig. 3, the insertion time costs of the two schemes increase with the used dataset size ratio. It takes 6,128 ms, 103,790 ms, and 383,615 ms for *SecGraph* to construct the encrypted database using the entire Email, Youtube, and Gplus datasets, respectively. It is worth noting that *SecGraph* is considerably up to $58\times$ faster than PeGraph. The reason is that PeGraph needs to perform a large number of expensive exponentiation modulo operations to calculate *xtag* for each keyword-identifier pair to build *XSet*, while *SecGraph* only computes the same number of hash values due to the design of the LDCF-encoded *XSet*. We further experiment that *SecGraph* takes on average 0.36 ms and 0.37 ms to insert and delete a pair, respectively. The experiment results demonstrate that has better update performance than PeGraph.

Our code: <https://github.com/XJTUOSV-SSEer/SecGraph>.

The code of LDCF: <https://github.com/CGCL-codes/LDCF>.

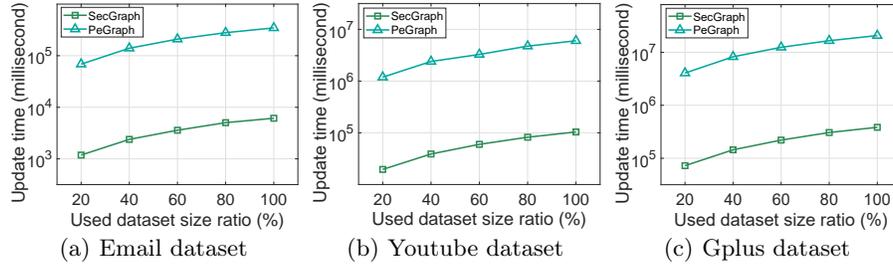


Fig. 3. Insertion performance in distinct datasets.

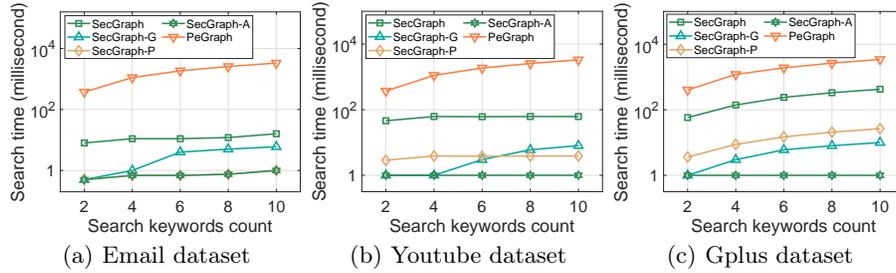


Fig. 4. Search performance in distinct datasets.

Search Performance. Here, we evaluate the search performance of all considered schemes as the search keyword count increases from 2 to 10. Fig. 4 depicts that, for every dataset, *SecGraph* takes less time to search compared with PeGraph and our optimized schemes perform better than *SecGraph*. Specifically, *SecGraph* yields up to $208\times$ improvement in search time compared with PeGraph, and *SecGraph-G*, *SecGraph-P* and *SecGraph-A* respectively yield up to $572\times$, $3,331\times$ and $3,430\times$ improvements in search time compared with PeGraph. First, there are two reasons why the search procedure in *SecGraph* is faster than that in PeGraph: (1) *SecGraph* only requires one roundtrip to send search tokens to the server due to the design of proxy-token generation method, while PeGraph requires two roundtrips, causing the communication cost in PeGraph is up to $540\times$ larger than that in *SecGraph*. (2) *SecGraph* generates one *xtag* just by computing a hash value to check whether or not it exists due to the design of the LDCF-encoded *XSet*, while PeGraph needs to execute an expensive exponentiation modulo operation to calculate a *xtag*. Then, *SecGraph-G* performs better than *SecGraph* since the former adopting the fingerprint grouping strategy reduces the number of *ocall* (i.e., the number of loaded *sub-filters*) by 96% compared with the latter. *SecGraph-P* performs better than *SecGraph* since the former parallelizes the procedure of the membership check of the *xtag*.

Furthermore, we evaluate the enclave storage cost of *SecGraph* and *SecGraph-G* as the search keywords count increases from 2 to 10. As shown in Fig.5, we observe different results on datasets of different scales. For the small-size dataset (i.e., Email dataset), the enclave storage cost in *SecGraph* hardly changes with the growth of search keywords count as there are only 20 *sub-filters* in total and when the search keywords count is 2, all *sub-filters* need to be loaded into the enclave. However, *SecGraph-G* only requires loading 1 *sub-filter* when the search keywords count is 2. For the middle-size dataset (i.e., Youtube dataset),

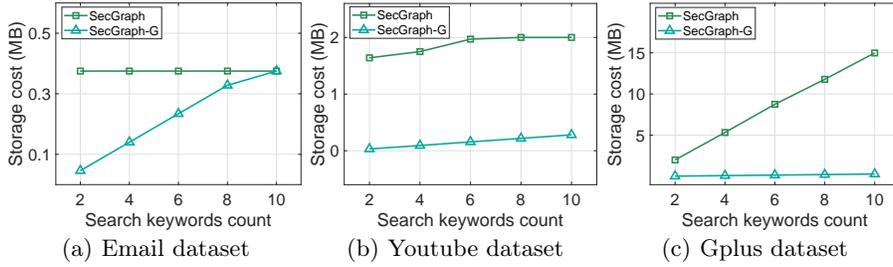


Fig. 5. Enclave storage cost in distinct datasets.

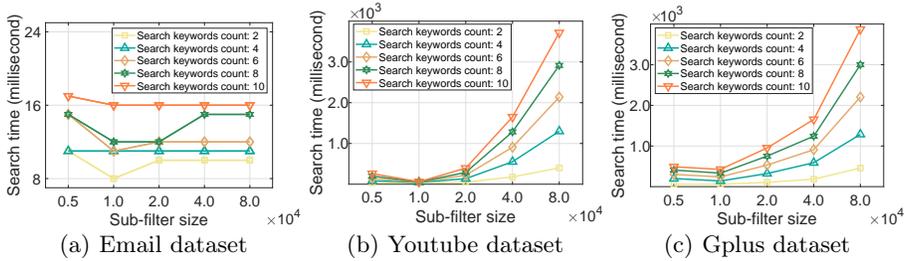


Fig. 6. Effect of the *sub-filter* size.

the enclave storage cost in *SecGraph* initially increases linearly with the search keywords count, until the search keywords count reaches 6, and the enclave storage cost in *SecGraph* remains stable and unchanged. The reason is that when the search keywords count is 6, all *sub-filters* need to be loaded into the enclave. But for *SecGraph-G*, the enclave storage cost increases linearly with the search keywords count. For the large-size dataset (i.e., Gplus dataset), the enclave storage cost in *SecGraph-G* also increases linearly with the search keywords count and there are 1408 *sub-filters* in total and even when the search keywords count is 10, only 9 *sub-filters* need to be loaded. Besides that, we surprisingly find that it only takes only 15 MB to store the loaded *sub-filters* in the enclave.

Effect of Parameters. Next, we evaluate the search time cost in *SecGraph* as the *sub-filter* size (i.e., the number of fingerprints contained in a *sub-filter*) increases from 5,000 to 80,000 under different search keywords count settings. Fig.6 shows that, for every dataset, the search time cost is the lowest when the *sub-filter* size is 10,000, which demonstrates the rationality of our default *sub-filter* size setting. We analyze that setting the *sub-filter* size too small will lead to a large number of *sub-filters* being loaded during a search procedure, resulting in a high number of *ocalls* and a long *ocall* time. However, setting it too large can reduce the number of *sub-filters* to be loaded appropriately, but it increases the amount of transferred data and also increases the *ocall* time.

Performance of Fuzzy Search Enabled *SecGraph*. Finally, we evaluate the performance of extended *SecGraph* that supports fuzzy search as the sub-string length increases from 2 bytes to 6 bytes in the Email dataset. Specifically, as shown in Fig. 7(a), the total insertion time cost in *SecGraph* decreases as the sub-string length increases. This is reasonable since the shorter the sub-string length, the more sub-strings each vertex’s *name* is split into, which also means more keywords to be extracted, requiring more time to construct the encrypted database. Fig. 7(b) shows that the storage cost of *TSet* decreases linearly as the sub-string length increases due to the same reason mentioned above. Fig. 7(c)

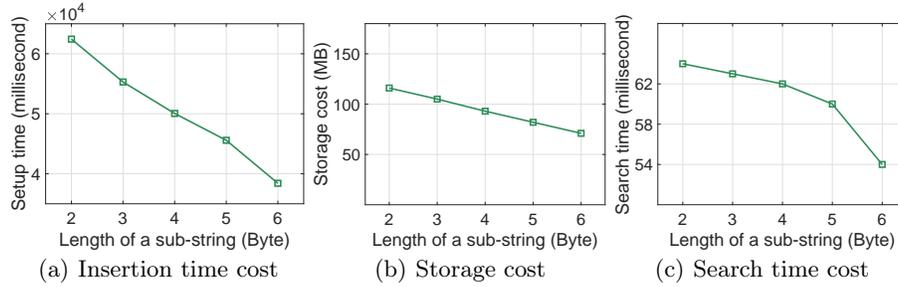


Fig. 7. Performance of fuzzy search enabled *SecGraph*.

shows that the search time slowly decreases as the sub-string length increases because the longer the sub-string length, the fewer matching results returned.

9 Conclusion

In this paper, we propose an SGX-based efficient and confidentiality-preserving graph search scheme *SecGraph* to provide secure and efficient search services over encrypted graphs. Firstly, we design a new proxy-token generation method to reduce the communication cost. Then, we design an LDCF-encoded *XSet* to reduce the computation cost. Moreover, we design a *Twin-TSet* to enable encrypted search over dynamic graphs. We further extend *SecGraph* into two optimized schemes *SecGraph-G* and *SecGraph-P* by adopting the fingerprint grouping and checking parallelization strategies, respectively, to speed up the search procedure. Finally, experiments and security analysis show that *SecGraph* can achieve secure and efficient search over dynamic graphs. In the future, we will explore efficient and confidentiality-preserving graph search schemes that support more plentiful search services such as range search, boolean search, etc.

References

1. Facebook. <http://www.facebook.com>
2. Instagram. <https://www.instagram.com/>
3. Cash, D., Jaeger, J., Jarecki, S., Jutla, C.S., Krawczyk, H., Rosu, M., Steiner, M.: Dynamic searchable encryption in very-large databases: Data structures and implementation. In: NDSS (2014)
4. Cash, D., Jarecki, S., Jutla, C.S., Krawczyk, H., Rosu, M., Steiner, M.: Highly-scalable searchable symmetric encryption with support for boolean queries. In: CRYPTO. pp. 353–373 (2013)
5. Chase, M., Kamara, S.: Structured encryption and controlled disclosure. In: ASIACRYPT. pp. 577–594 (2010)
6. Chen, X., Huo, H., Huan, J., Vitter, J.S., Zheng, W., Zou, L.: Msq-index: A succinct index for fast graph similarity search. IEEE TKDE. **33**(6), 2654–2668 (2021)
7. Costan, V., Devadas, S.: Intel sgx explained. In: Cryptology ePrint Archive (2016)
8. Demmler, D., Schneider, T., Zohner, M.: ABY - A framework for efficient mixed-protocol secure two-party computation. In: NDSS (2015)
9. Fan, B., Andersen, D.G., Kaminsky, M., Mitzenmacher, M.: Cuckoo filter: Practically better than bloom. In: CoNEXT. pp. 75–88. ACM (2014)

10. Gao, C., Wang, X., He, X., Li, Y.: Graph neural networks for recommender system. In: WSDM. pp. 1623–1625 (2022)
11. Ghosh, E., Kamara, S., Tamassia, R.: Efficient graph encryption scheme for shortest path queries. In: CCS. pp. 516–525 (2021)
12. Lai, S., Yuan, X., Sun, S., Liu, J.K., Liu, Y., Liu, D.: Graphse²: An encrypted graph database for privacy-preserving social search. In: AsiaCCS. pp. 41–54 (2019)
13. Patrnanabis, S., Mukhopadhyay, D.: Forward and backward private conjunctive searchable symmetric encryption. In: NDSS (2021)
14. Shen, M., Wang, M., Xu, K., Zhu, L.: Privacy-preserving approximate top-k nearest keyword queries over encrypted graphs. In: IWQOS. pp. 1–10 (2021)
15. Wang, H., Zhao, Z., Yang, K., Song, H., Xiao, Y.: Approximate nearest neighbor search using query-directed dense graph. In: DASFAA Workshops. pp. 429–444 (2021)
16. Wang, S., Zheng, Y., Jia, X., Yi, X.: Pegraph: A system for privacy-preserving and efficient search over encrypted social graphs. *IEEE TIFS*. **17**, 3179–3194 (2022)
17. Zhang, F., Chen, H., Jin, H., Reviriego, P.: The logarithmic dynamic cuckoo filter. In: *IEEE ICDE*. pp. 948–959 (2021)
18. Zhang, K., Tong, Y., Shi, Y., Zeng, Y., Xu, Y., Chen, L., Zhou, Z., Xu, K., Lv, W., Zheng, Z.: Approximate k-nearest neighbor query over spatial data federation. In: *DAFSAA*. pp. 351–368 (2023)
19. Zhao, Z., Qian, P., Yang, X., Zeng, Z., Guan, C., Leong, T.W., Li, X.: Semignn-ppi: Self-ensembling multi-graph neural network for efficient and generalizable protein-protein interaction prediction. In: *IJCAI*. pp. 4984–4992 (2023)