Online Resource 1

1 Literature Overview

1.1 Symmetric Searchable Encryption (SSE)

The notion of keyword search across encrypted data was introduced by Song et al. [1]. In their work, authors presented an encryption algorithm where each input document is uploaded onto cloud storage server in terms of encrypted keywords. Subsequently, data user asks server to search across document by issuing a search token. Though the scheme is practical, it incurs search overhead linear to total keywords for all documents i.e. $O(\sum_{1\leq i\leq m}|D_i|)$. In addition, the scheme [1] reveals information about keywords to be searched to server. The issue was later addressed by Goh et al. [2] where a secure index (list of encrypted keywords) is associated with a ciphertext of original document. Additionally, Goh et al. defined the first formal security model for searchable encryption in [2]. Subsequently, several schemes discussed in [3–7] with the improved security and efficiency have been proposed. However, schemes in [1–7] focus on a single keyword search and return several irrelevant results that ultimately consume more resources. To address this issue, numerous schemes for multi-keyword search have been proposed with either conjunctive keyword search [8–13] or boolean keyword search [14,15] functionality.

From literature, we observe that most SSE schemes [3–7, 12, 14] use II search structure where a secure common index prepared from the available collection of documents and set of keywords, is uploaded onto server along with actual document ciphertexts. However, such II search structure would lead to inefficient search operation in case of multi-keyword SSE [12,14]. The reason behind such inefficiency is the usage of either of the approaches i.e. set intersection operation at server (that leaks keyword-data relationship) or storage of $O(2^V)$ meta-keywords at server for V keywords (that introduces exponential storage overhead) [8,9]. More precisely, with II search structure, the server-side storage overhead for both the schemes [12,14] is $O(|D| + V + \sum|D_w|)$ for $|D|$ number of documents, V number of keywords and $|D_w|$ number of documents containing keyword $w \in V$ (Note that V is defined in Table 1). In addition, the size of token in [14] is linear to the size of query i.e. $O(|Q|)$. On the other hand, though the scheme [12] with conjunctive search offers lesser query expressiveness than boolean search of [14], it provides constant sized token. Recently, Hu et al. offers two II-based SSE schemes [13] using bloom filter and Inner Product Encryption (IPE) respectively. Besides conjunctive keyword search, both the schemes provide forward secure SSE where document insertion does not leak information about keywords contained in that document. However, the first scheme may have false positive problem - an inherent issue of bloom filter. On the other hand, though the second scheme avoids such problem, it suffers from high computational complexity due to the usage of IPE mechanism where complexity is $O(V)$. Contrast to II-based SSE schemes, the SI-based schemes could offer multi-keyword search without any leakage or additional storage overhead onto server. However, in literature a very few schemes exist with SI search structure [8–11].

1.2 SSE with dynamic index update

The initial SSE schemes [3,4,7,8,14,16] work for the static data collection. However, support of dynamic data update is indeed essential in real world applications. As mentioned previously, in SI-based SSE schemes, each document has its own index containing keyword fields and thus insertion of a new document in collection does not affect the other documents or their indexes. Correspondingly, update (insert/delete of a document) in data collection does not require any index update operation in such SI-based SSE schemes [8–11]. However, most SSE schemes have been defined based on II-search structure

and so index update becomes a crucial research topic since the last decade. Several works concerning dynamic SSE have been proposed in [5, 6, 17–24]. Kamara et al. have defined the first dynamic SSE scheme [5] where insertion/deletion of a document is performed by updating a table-based II. The scheme provides adaptive security against chosen keyword attack and sublinear search time. Furthermore, an advanced dynamic SSE for multi-processor system has been proposed in [6] using red-black tree-based II. However, the size of II in both the schemes [5,6] is $O((V+P|D_w|)$. Cash et al. [14] have proposed dynamic SSE for large data collection with the reduced storage overhead i.e. $O(P|D_w|)$ at server. Furthermore, the schemes [18, 21] allow document insertion /deletion/modification by updating matrix-based II. The schemes [19, 22] offer dynamic SSE for link-list based II. In such schemes, document insertion/deletion needs modification in search list, search array and corresponding link list as well as delete table (in case of deletion). However, document modification (insertion of keyword(s)) is not supported by [19, 22]. The generalized update complexity for the schemes [18, 19, 21, 22] is $O(V \cdot |D|)$. Recently, Ge et al. [20] have offered dynamic SSE with the list and array based inverted index as that in [19, 22] with the improved index update complexity of $O(V)$. On the other hand, Xia et al. [23] offers dynamic SSE for tree-based II with update complexity $O(V^2 log|D|)$. Note that for all dynamic SSE schemes, the user is responsible to compute update token of size linear to the number of keywords included by the document to be inserted or deleted.

1.3 SSE with result verification

Majority of SSE schemes assume that the cloud server is semi-honest-but-curious that honestly executes the search operation across the available ciphertexts and sends the search result intact to the requesting data user. From the search result, the server only attempts to learn plaintext [25]. However, in practice, the search result may be corrupted due to hardware/software failures. Moreover, even though the server executes the protocol honestly, it may be attacked by a dishonest user who may direct the server to modify the search result. Ultimately, the server behaves maliciously and outputs an altered or tampered search result. Thus, it is indeed essential for data user to check the correctness of the available search result. In order to attain such a requirement, several SSE schemes [3, 18–20, 25–35] have been defined with provision of query result verification at user side. Amongst them, the schemes in [28–30] support relational databases and the corresponding query operations (viz. Projection, Selection, Join etc.). The other schemes [3, 18–20, 25–27, 31–35] support keyword based search queries for the unstructured dataset (i.e. document collection). Note that in all verification enabled SSE schemes, besides collection of encrypted documents and search index, a server-side verification component is uploaded onto server. The server after performing search operation computes a proof component from such a verification component. More precisely, the scheme [18, 19, 25–27, 31–33] offers result verification using tree-based verification component. Amongst them, the schemes [25–27, 31–33] support a single-keyword search only. Additionally, in all these schemes, the proof component includes subtree computed from the original tree and thus, the size of proof component is of significant overhead for these schemes. Furthermore, user-side result verification requires tree reconstruction that would be computationally expensive in case of large number of keywords in system. On the other hand, the state-of-the-art works given in [18,19] offer II-based verification enabled SSE with dynamic data update. Both schemes use bilinear map accumulator based verification tree and offers multi-keyword search with set-intersection operation at server side. A proof component in [18,19] includes accumulation values for all nodes from the search keywords to the root node of tree. Thus, the size of proof component and its computational complexity is linear to query size. Another multi-keyword ranked searchable encryption scheme has been proposed by Jiang et al. [35] where a special data structure known as QSet is uploaded onto server as a verification component. With QSet, this scheme offers multi-keyword search with sublinear search time. A more efficient and privacy preserving multi-keyword scheme with result verification is proposed by Wan et al. [34] where adapted homomorphic MAC based authentication tag is used as a server-side verification component. With such tag, cloud server can homomorphically execute the search function over the authentication tags to derive the result with a proof, which can certify the search result. Recently, Ge et al. have proposed II-based multi-keyword SSE [20] with result verification based on a novel approach of Accumulative Authentication Tag (AAT). In this scheme, a Verification List (VL) which is a link list of nodes for each keywords is uploaded onto server as a server-side verification component. Each node in VL includes AAT component computed from the set of documents containing the corresponding

keywords. With AAT, the scheme assures security against collision attack (that is identifying different messages with the same tag) and replay attack (that is sending old data in result even though data has been updated). Moreover, the scheme offers dynamic index update operation.

1.4 ASE with result verification

In asymmetric key setting, there do exist few searchable encryption schemes offering search result verification. The scheme [36] provides public delegatability (that is, any user - not necessarily the data owner, could issue a search query to the server) and public verifiability (that is, any third party data user not necessarily the user initiating the search query, by possessing public verifiable key, would be able to verify the search result) [37][48]. However, the scheme in [36] offers search operation across plaintext outsourced database. On the other hand, public verifiability for the encrypted data has been proposed by Cheng et al. [33]. The scheme offers indistinguishability obfuscation based verification circuit that allows any user possessing public verification key to verify the search result. However, generation and obfuscation of such verification circuit entails substantial cost linear to $O(V \cdot |D|)$. Furthermore, in [38] Zheng et al. proposed the notion of VABKS- Verifiable Attribute Base Keyword Search, utilizing fine-grained access control mechanism. The data owner in VABKS allows a data user satisfying the access control policy (associated with each keyword) to issue search query and later verify the search result. However, the scheme assumes existence of a secure channel amongst the communicating entities. A secure channel free, access control based, public keyword searchable encryption scheme with result verifiability has been proposed by Liu et al. [39]. Though such public-key based keyword search mechanisms offer result verification, they support the static data collection and fixed set of keywords due to utilization of Inverted Index search structure in data encryption. However the focus of our research discussed in this paper is to illustrate the benefits of using Simple Index search structure over the Inverted Index search structure and potential mechanism of result verification in SI-based SSE schemes. Therefore, we consider the schemes using result verification in asymmetric key searchable encryption setting, as out of scope.

References

[1] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*, pages 44–55. IEEE, 2000.

[2] Eu-Jin Goh et al. Secure indexes. *IACR Cryptology ePrint Archive*, 2003:216, 2003.

[3] Kaoru Kurosawa and Yasuhiro Ohtaki. Uc-secure searchable symmetric encryption. In *International Conference on Financial Cryptography and Data Security*, pages 285–298. Springer, 2012.

[4] Reza Curtmola and Ostrovsky Garay, Kamara Seny. Searchable symmetric encryption: improved definitions and efficient constructions. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 79–88. ACM, 2006.

[5] Seny Kamara, Charalampos Papamanthou, and Tom Roeder. Dynamic searchable symmetric encryption. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 965–976. ACM, 2012.

[6] Seny Kamara and Charalampos Papamanthou. Parallel and dynamic searchable symmetric encryption. In *Financial Cryptography and Data Security*, pages 258–274. Springer, 2013.

[7] Peter Van Liesdonk, Saeed Sedghi, Jeroen Doumen, Pieter Hartel, and Willem Jonker. Computationally efficient searchable symmetric encryption. In *Secure data management*, pages 87–100. Springer, 2010.

[8] Philippe Golle, Jessica Staddon, and Brent Waters. Secure conjunctive keyword search over encrypted data. In *Applied Cryptography and Network Security*, pages 31–45. Springer, 2004.

[9] Lucas Ballard, Seny Kamara, and Fabian Monrose. Achieving efficient conjunctive keyword searches over encrypted data. In *Information and Communications Security*, pages 414–426. Springer, 2005.

[10] Jin Wook Byun, Dong Hoon Lee, and Jongin Lim. Efficient conjunctive keyword search on encrypted data storage system. In *European Public Key Infrastructure Workshop*, pages 184–196. Springer, 2006.

[11] Peishun Wang, Huaxiong Wang, and Josef Pieprzyk. Keyword field-free conjunctive keyword searches on encrypted data and extension for dynamic groups. In *Cryptology and Network Security*, pages 178–195. Springer, 2008.

[12] Haiping Huang, Jianpeng Du, Hui Wang, and Ruchuan Wang. A multi-keyword multi-user searchable encryption scheme based on cloud storage. In *Trustcom/BigDataSE/I SPA, 2016 IEEE*, pages 1937–1943. IEEE, 2016.

[13] Chengyu Hu, Xiangfu Song, Pengtao Liu, Yue Xin, Yuqin Xu, Yuyu Duan, and Rong Hao. Forward secure conjunctive-keyword searchable encryption. *IEEE Access*, 7:35035–35048, 2019.

[14] David Cash, Stanislaw Jarecki, Charanjit Jutla, Hugo Krawczyk, Marcel-Cătălin Roşu, and Michael Steiner. Highly-scalable searchable symmetric encryption with support for boolean queries. In *Advances in Cryptology–CRYPTO 2013*, pages 353–373. Springer, 2013.

[15] Tarik Moataz and Abdullatif Shikfa. Boolean symmetric searchable encryption. In *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, pages 265–276. ACM, 2013.

[16] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on parallel and distributed systems*, 25(1):222–233, 2014.

[17] David Cash, Joseph Jaeger, Stanislaw Jarecki, Charanjit Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Dynamic searchable encryption in very-large databases: Data structures and implementation. In *Network and Distributed System Security Symposium (NDSS14)*, 2014.

[18] Wenhai Sun, Xuefeng Liu, Wenjing Lou, Y Thomas Hou, and Hui Li. Catch you if you lie to me: Efficient verifiable conjunctive keyword search over large dynamic encrypted cloud data. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2110–2118. IEEE, 2015.

[19] Yuxi Li, Fucai Zhou, Yuhai Qin, Muqing Lin, and Zifeng Xu. Integrity-verifiable conjunctive keyword searchable encryption in cloud storage. *International Journal of Information Security*, 17(5):549–568, 2018.

[20] Xinrui Ge, Jia Yu, Hanlin Zhang, Chengyu Hu, Zengpeng Li, Zhan Qin, and Rong Hao. Towards achieving keyword search over dynamic encrypted cloud data with symmetric-key based verification. *IEEE Transactions on Dependable and Secure Computing*, 2019.

[21] Xiaoyu Zhu, Qin Liu, and Guojun Wang. A novel verifiable and dynamic fuzzy keyword search scheme over encrypted data in cloud computing. In *2016 IEEE Trustcom/BigDataSE/ISPA*, pages 845–851. IEEE, 2016.

[22] Qin Liu, Xiaohong Nie, Xuhui Liu, Tao Peng, and Jie Wu. Verifiable ranked search over dynamic encrypted data in cloud computing. In *Quality of Service (IWQoS), 2017 IEEE/ACM 25th International Symposium on*, pages 1–6. IEEE, 2017.

[23] Zhihua Xia, Xinhui Wang, Xingming Sun, and Qian Wang. A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. *IEEE transactions on parallel and distributed systems*, 27(2):340–352, 2015.

[24] Cheng Guo, Xue Chen, Yingmo Jie, Fu Zhangjie, Mingchu Li, and Bin Feng. Dynamic multi-phrase ranked search over encrypted data with symmetric searchable encryption. *IEEE Transactions on Services Computing*, 2017.

[25] Qi Chai and Guang Gong. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers. In *Communications (ICC), 2012 IEEE International Conference on*, pages 917–922. IEEE, 2012.

[26] Jianfeng Wang, Xiaofeng Chen, Hua Ma, Qiang Tang, Jin Li, and Hui Zhu. A verifiable fuzzy keyword search scheme over encrypted data. *Journal of Internet Services and Information Security (JISIS)*, 2:49–58, 2012.

[27] HweeHwa Pang and Kyriakos Mouratidis. Authenticating the query results of text search engines. *Proceedings of the VLDB Endowment*, 1(1):126–137, 2008.

[28] HweeHwa Pang and K-L Tan. Authenticating query results in edge computing. In *Data Engineering, 2004. Proceedings. 20th International Conference on*, pages 560–571. IEEE, 2004.

[29] Feifei Li, Marios Hadjieleftheriou, George Kollios, and Leonid Reyzin. Dynamic authenticated index structures for outsourced databases. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 121–132. ACM, 2006.

[30] Yanbin Lu. Privacy preserving logarithmic-time search on encrypted data in cloud. In *NDSS*, 2012.

[31] Zachary A Kissel and Jie Wang. Verifiable phrase search over encrypted data secure against a semi-honest-but-curious adversary. In *Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on*, pages 126–131. IEEE, 2013.

[32] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y Thomas Hou, and Hui Li. Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking. *IEEE Transactions on Parallel and Distributed Systems*, 25(11):3025–3035, 2014.

[33] Rong Cheng, Jingbo Yan, Chaowen Guan, Fangguo Zhang, and Kui Ren. Verifiable searchable symmetric encryption from indistinguishability obfuscation. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, pages 621–626. ACM, 2015.

[34] Zhiguo Wan and Robert H Deng. Vpsearch: Achieving verifiability for privacy-preserving multi-keyword search over encrypted cloud data. *IEEE Transactions on Dependable and Secure Computing*, 2016.

[35] Xiuxiu Jiang, Jia Yu, Jingbo Yan, and Rong Hao. Enabling efficient and verifiable multi-keyword ranked search over encrypted cloud data. *Information Sciences*, 403:22–41, 2017.

[36] Monir Azraoui, Kaoutar Elkhiyaoui, Melek Önen, and Refik Molva. Publicly verifiable conjunctive keyword search in outsourced databases. In *Communications and Network Security (CNS), 2015 IEEE Conference on*, pages 619–627. IEEE, 2015.

[37] Bryan Parno, Mariana Raykova, and Vinod Vaikuntanathan. How to delegate and verify in public: Verifiable computation from attribute-based encryption. In *Theory of Cryptography Conference*, pages 422–439. Springer, 2012.

[38] Qingji Zheng, Shouhuai Xu, and Giuseppe Ateniese. Vabks: verifiable attribute-based keyword search over outsourced encrypted data. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 522–530. IEEE, 2014.

[39] Pengliang Liu, Jianfneg Wang, Hua Ma, and Haixin Nie. Efficient verifiable public key encryption with keyword search based on kp-abe. In *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2014 Ninth International Conference on*, pages 584–589. IEEE, 2014.