1 Introduction

2 Related Work

In this section, we give a brief account of constructions of searchable encryption scheme, known attacks on them, and counter-measures proposed to negate the attacks. We argue that the current counter-measures are insufficient to protect the underlying data.

2.1 Keyword Search in Encrypted Databases

There is a wide range of schemes proposed to allow for keyword search in encrypted databases [14].

Schemes which use symmetric key primitives are known as SSE schemes. Song et al. [38] proposed the first SSE scheme. The construction essentially puts keywords into a chain and encrypt it that way, such that the whole chain has to be visited when a search query is issued. The scheme is proven to be a secure encryption scheme but it is not proven to be a secure encryption scheme. In fact, frequency of the underlying plaintexts are revealed upon searching so the scheme is vulnerable to statistical attacks such as inference analysis. More recent works such as [16, 11, 12, 21, 39, 8, 33, 18, 24, 15, 23] proposes to build encrypted inverted index for the keywords in the document collection. The time complexity of search query is proportional to the number of keywords in the database. In terms of security notion, Goh [16] proposed semantic security against adaptive chosen keyword attack (IND-CKA and IND2-CKA), and Chang and Mitzenmacher [11] introduced a simulation-based definition that is slightly stronger. Both of the definitions are shown to be insufficient in [12]. The authors in [12] has also formulated new security notions both in non-adaptive setting and in adaptive setting.

On the other hand, schemes which use public key primitives are known as PEKS schemes. The first PEKS scheme was proposed by Boneh et al. [3]. Later on, Bellare et al. [2] introduced the notion of public key efficiently searchable encryption (ESE) and proposed constructions in the random oracle model. The main difference between public key ESE and symmetric key solutions is that public key ESE uses deterministic encryption on keywords so the frequency of keywords are leaked before any queries. Subsequent work focus on enhancing the security notion of ESE and to protect the database from keyword guessing attack.

Recently, there has been an extensive work attempting to enrich expressiveness of the search function. Fuzzy keyword search enhances functionality by allowing non-exact matches. For example, search of 'rael' returns documents containing 'real' too. Fuzzy keyword search has been studied in symmetric key setting in [30, 29, 42, 44] and in public key setting in [6, 42]. Conjunctive keyword search allows the user to search for multiple keywords. It has been studied in symmetric key settings in [17, 1, 41, 9, 13] and in public key setting in [35, 4, 37, 19, 27, 22, 26]. It is possible for the server to be dishonest and return only partial or inaccurate results. Verifiable keyword search is devised to tackle the problem. In symmetric key setting [34, 28, 10, 40] are proposed. In public key setting [43, 31] are proposed.

2.2 Attacks on Current Schemes

There are many known genres of attacks on searchable encryption. In this paper, we focus on the attacks abusing keyword frequencies in the database. Some of the schemes leak frequencies before any queries, while some other schemes only leak frequencies of the queried keywords. We stress that the leakage profile in the latter case is not good enough to protect the database against statistical attack, since with enough queries, the adversary can recover frequency of all keywords.

The first frequency-based attack on searchable encryption, known as IKK attack, is presented by Islam et

al. [20]. In their attack model, the adversary is assumed to know the distribution of number of documents of pairs of keywords and his goal is to recover trapdoors used in queries. By comparing the distribution to the observed counts in queries using simulated annealing, the authors were able to recover almost all queries for relatively small ($\sim$500) keyword set size, and for keyword set size of 2500, their attack could successfully identify most of the queries. The efficiency and accuracy of the attack is later improved by Pouliot et al. [36] by casting the problem as a graph matching problem.

By observing that a large fraction of keywords match to unique number of documents, Cash et al. proposed count attack [7]. Their attack achieves better recovery rate of queries compared to IKK attack. Naveed et al. [32] have demonstrated that even simple frequency analysis is sufficient to recover encrypted keywords in databases.

TODO: one more attack proposed in 2017

2.3 Countermeasures

Islam et al. [20] suggested $(\alpha, t)-$secure index to prevent inference attack. The scheme aims to have partitions of keywords of size at least α each, such that in each partition, the sets of documents containing those keywords differ by at most t documents. They proposed an algorithm to achieve $(\alpha, 0)$-secure index for any databases. The scheme uses clustering algorithm on the database so it is static in nature.

Lacharité et al. proposed to use frequency-smoothing encryption to defend against frequency analysis [25]. The authors have only considered single-column encryption in relational databases so the scheme is not provably secure against attacks based on co-occurrence of keywords. Bost et al. [5] attempted to fix security notion for searchable encryption by posting additional constraints. They proposed a padding-based scheme and tested the scheme against count attack [7]. The experiment results show that their scheme is still vulnerable to count attack.

To our best knowledge, there is yet a satisfactory security notion for searchable encryption that is secure against statistical attacks (leakage-abuse attacks in general), and there is no scheme known to be secure under all statistical attacks.

3 Notation

Notation	**Explanation**
DB	Database that has not been encrypted
EDB	Encrypted database
d	A document in the database
f	A file associated to some document
K	A set of keywords associated to some file
k	A keyword in a keyword set
ed	An encrypted document
ef	An encrypted file
EK	A set of encrypted keywords
ek	An encrypted keyword

Table 1: Notations used in this paper.

4 Security Notions

5 Construction

5.1 Motivation

Sort this out only when we settle on security notion because they are supposed to be closely related.

5.2 Permutable Matrix

In this section, we will define permutable matrix. In short, permutable matrix is a special type of matrix with the property that after applying row and column permutations, we get the same matrix as before. After that, we generalize the idea to higher dimensions.

Definition 1 *Let M be a two-dimensional n by n matrix. Let P, Q be permutation matrices on $[n]$. Then we say M is (P, Q)-permutable if $PMQ^T = M$. If $P = Q$, we simply say M is P-permutable.*

We stress that the following is not a proof of security. Imagine we have G an co-occurrence count matrix on keywords $(m_1, \cdots m_k)$. The adversary knows G and a permutation of it, say $H = QGQ^T$ for some permutation matrix Q. His goal is to recover Q so he can map the encrypted keywords to the plaintexts. Indeed, we know from [36] that this can be done with high accuracy. We imagine the worst case where the adversary recovers Q completely. If Q is the unique solution to $\min_X \left\|XGX^T - H\right\|$, then the adversary can recover all keywords with certainty. However, if G is P-permutable, we see $\left\|QGQ^T - H\right\| = \left\|QPGP^TQ^T - H\right\| = \left\|(QP)G(QP)^T - H\right\|$, so QP also minimizes the objective function. In fact, all permutations of the shape QP^i for $i = 0, \cdots, k-1$ minimizes the norm. In that case, the adversary can no longer map keywords to their encryptions uniquely. Since P is a permutation of full length, QP^i sends keywords to different encryptions for each i. Hence, if we start with an co-occurrence count matrix that permutable by some permutation of full length, we can significantly reduce the likelihood of the adversary recovering encryptions of keywords via co-occurrence based attacks.

Probably have to change after settle on security notion. The argument above is not a proof of security for three reasons. First of all, co-occurrence matrix itself is aggregated information, we cannot assume that it is the only information the adversary uses. Secondly, there is no reason to assume that co-occurrence matrix based attack is the only technique that can be used to breach security. Finally, the goal of the adversary is not necessarily keyword recovery. For instance, he can ask the following question: 'Is keyword k in either file f_0 or file f_1?'.

5.3 Permutation-based Padding

References

[1] Lucas Ballard, Seny Kamara, and Fabian Monrose. Achieving efficient conjunctive keyword searches over encrypted data. In Sihan Qing, Wenbo Mao, Javier López, and Guilin Wang, editors, *ICICS 05: 7th International Conference on Information and Communication Security*, volume 3783 of *Lecture Notes in Computer Science*, pages 414–426, Beijing, China, December 10–13, 2005. Springer, Heidelberg, Germany.

[2] Mihir Bellare, Alexandra Boldyreva, and Adam O'Neill. Deterministic and efficiently searchable encryption. In Alfred Menezes, editor, *Advances in Cryptology – CRYPTO 2007*, volume 4622 of *Lecture Notes in Computer Science*, pages 535–552, Santa Barbara, CA, USA, August 19–23, 2007. Springer, Heidelberg, Germany.

[3] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public key encryption with keyword search. In Christian Cachin and Jan Camenisch, editors, *Advances in Cryptology – EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 506–522, Interlaken, Switzerland, May 2–6, 2004. Springer, Heidelberg, Germany.

[4] Dan Boneh and Brent Waters. Conjunctive, subset, and range queries on encrypted data. In Salil P. Vadhan, editor, *TCC 2007: 4th Theory of Cryptography Conference*, volume 4392 of *Lecture Notes in Computer Science*, pages 535–554, Amsterdam, The Netherlands, February 21–24, 2007. Springer, Heidelberg, Germany.

[5] Raphael Bost and Pierre-Alain Fouque. Thwarting leakage abuse attacks against searchable encryption – A formal approach and applications to database padding. Cryptology ePrint Archive, Report 2017/1060, 2017. `http://eprint.iacr.org/2017/1060`.

[6] J. Bringer, H. Chabanne, and B. Kindarji. Error-tolerant searchable encryption. In *2009 IEEE International Conference on Communications*, pages 1–6, June 2009.

[7] David Cash, Paul Grubbs, Jason Perry, and Thomas Ristenpart. Leakage-abuse attacks against searchable encryption. In Indrajit Ray, Ninghui Li, and Christopher Kruegel:, editors, *ACM CCS 15: 22nd Conference on Computer and Communications Security*, pages 668–679, Denver, CO, USA, October 12–16, 2015. ACM Press.

[8] David Cash, Joseph Jaeger, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Dynamic searchable encryption in very-large databases: Data structures and implementation. In *ISOC Network and Distributed System Security Symposium – NDSS 2014*, San Diego, CA, USA, February 23–26, 2014. The Internet Society.

[9] David Cash, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Highly-scalable searchable symmetric encryption with support for Boolean queries. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013, Part I*, volume 8042 of *Lecture Notes in Computer Science*, pages 353–373, Santa Barbara, CA, USA, August 18–22, 2013. Springer, Heidelberg, Germany.

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

[11] Yan-Cheng Chang and Michael Mitzenmacher. Privacy preserving keyword searches on remote encrypted data. Cryptology ePrint Archive, Report 2004/051, 2004. `http://eprint.iacr.org/2004/051`.

[12] Reza Curtmola, Juan A. Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors, *ACM CCS 06: 13th Conference on Computer and Communications Security*, pages 79–88, Alexandria, Virginia, USA, October 30 – November 3, 2006. ACM Press.

[13] Sky Faber, Stanislaw Jarecki, Hugo Krawczyk, Quan Nguyen, Marcel-Catalin Rosu, and Michael Steiner. Rich queries on encrypted data: Beyond exact matches. In Günther Pernul, Peter Y. A. Ryan, and Edgar R. Weippl, editors, *ESORICS 2015: 20th European Symposium on Research in Computer Security, Part II*, volume 9327 of *Lecture Notes in Computer Science*, pages 123–145, Vienna, Austria, September 21–25, 2015. Springer, Heidelberg, Germany.

[14] Benjamin Fuller, Mayank Varia, Arkady Yerukhimovich, Emily Shen, Ariel Hamlin, Vijay Gadepally, Richard Shay, John Darby Mitchell, and Robert K. Cunningham. SoK: Cryptographically protected database search. In *2017 IEEE Symposium on Security and Privacy*, pages 172–191, San Jose, CA, USA, May 22–26, 2017. IEEE Computer Society Press.

[15] Sebastian Gajek. Dynamic symmetric searchable encryption from constrained functional encryption. In Kazue Sako, editor, *Topics in Cryptology – CT-RSA 2016*, volume 9610 of *Lecture Notes in Computer Science*, pages 75–89, San Francisco, CA, USA, February 29 – March 4, 2016. Springer, Heidelberg, Germany.

[16] Eu-Jin Goh. Secure indexes. Cryptology ePrint Archive, Report 2003/216, 2003. `http://eprint.iacr.org/2003/216`.

[17] Philippe Golle, Jessica Staddon, and Brent R. Waters. Secure conjunctive keyword search over encrypted data. In Markus Jakobsson, Moti Yung, and Jianying Zhou, editors, *ACNS 04: 2nd International Conference on Applied Cryptography and Network Security*, volume 3089 of *Lecture Notes in Computer Science*, pages 31–45, Yellow Mountain, China, June 8–11, 2004. Springer, Heidelberg, Germany.

[18] Florian Hahn and Florian Kerschbaum. Searchable encryption with secure and efficient updates. In Gail-Joon Ahn, Moti Yung, and Ninghui Li, editors, *ACM CCS 14: 21st Conference on Computer and Communications Security*, pages 310–320, Scottsdale, AZ, USA, November 3–7, 2014. ACM Press.

[19] Yong Ho Hwang and Pil Joong Lee. Public key encryption with conjunctive keyword search and its extension to a multi-user system. In Tsuyoshi Takagi, Tatsuaki Okamoto, Eiji Okamoto, and Takeshi Okamoto, editors, *PAIRING 2007: 1st International Conference on Pairing-based Cryptography*, volume 4575 of *Lecture Notes in Computer Science*, pages 2–22, Tokyo, Japan, July 2–4, 2007. Springer, Heidelberg, Germany.

[20] Mohammad Saiful Islam, Mehmet Kuzu, and Murat Kantarcioglu. Access pattern disclosure on searchable encryption: Ramification, attack and mitigation. In *ISOC Network and Distributed System Security Symposium – NDSS 2012*, San Diego, CA, USA, February 5–8, 2012. The Internet Society.

[21] Seny Kamara and Charalampos Papamanthou. Parallel and dynamic searchable symmetric encryption. In Ahmad-Reza Sadeghi, editor, *FC 2013: 17th International Conference on Financial Cryptography and Data Security*, volume 7859 of *Lecture Notes in Computer Science*, pages 258–274, Okinawa, Japan, April 1–5, 2013. Springer, Heidelberg, Germany.

[22] Jonathan Katz, Amit Sahai, and Brent Waters. Predicate encryption supporting disjunctions, polynomial equations, and inner products. *Journal of Cryptology*, 26(2):191–224, April 2013.

[23] Kee Sung Kim, Minkyu Kim, Dongsoo Lee, Je Hong Park, and Woo-Hwan Kim. Forward secure dynamic searchable symmetric encryption with efficient updates. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *ACM CCS 17: 24th Conference on Computer and Communications Security*, pages 1449–1463, Dallas, TX, USA, October 31 – November 2, 2017. ACM Press.

[24] Kaoru Kurosawa, Keisuke Sasaki, Kiyohiko Ohta, and Kazuki Yoneyama. UC-secure dynamic searchable symmetric encryption scheme. In Kazuto Ogawa and Katsunari Yoshioka, editors, *IWSEC 16: 11th International Workshop on Security, Advances in Information and Computer Security*, volume 9836 of *Lecture Notes in Computer Science*, pages 73–90, Tokyo, Japan, September 12–14, 2016. Springer, Heidelberg, Germany.

[25] Marie-Sarah Lacharit and Kenneth G. Paterson. Frequency-smoothing encryption: preventing snapshot attacks on deterministically-encrypted data. Cryptology ePrint Archive, Report 2017/1068, 2017. `https://eprint.iacr.org/2017/1068`.

[26] Junzuo Lai, Xuhua Zhou, Robert Huijie Deng, Yingjiu Li, and Kefei Chen. Expressive search on encrypted data. In Kefei Chen, Qi Xie, Weidong Qiu, Ninghui Li, and Wen-Guey Tzeng, editors, *ASIACCS 13: 8th ACM Symposium on Information, Computer and Communications Security*, pages 243–252, Hangzhou, China, May 8–10, 2013. ACM Press.

[27] Allison B. Lewko, Tatsuaki Okamoto, Amit Sahai, Katsuyuki Takashima, and Brent Waters. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. In Henri Gilbert, editor, *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 62–91, French Riviera, May 30 – June 3, 2010. Springer, Heidelberg, Germany.

[28] Feifei Li, Marios Hadjieftheriou, George Kollios, and Leonid Reyzin. *Authenticated Index Structures for Outsourced Databases*, pages 115–136. Springer US, Boston, MA, 2008.

[29] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou. Fuzzy keyword search over encrypted data in cloud computing. In *2010 Proceedings IEEE INFOCOM*, pages 1–5, March 2010.

[30] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Enabling efficient fuzzy keyword search over encrypted data in cloud computing. Cryptology ePrint Archive, Report 2009/593, 2009. `http://eprint.iacr.org/2009/593`.

[31] P. Liu, J. Wang, H. Ma, and H. Nie. Efficient verifiable public key encryption with keyword search based on kp-abe. In *2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications*, pages 584–589, Nov 2014.

[32] Muhammad Naveed, Seny Kamara, and Charles V. Wright. Inference attacks on property-preserving encrypted databases. In Indrajit Ray, Ninghui Li, and Christopher Kruegel:, editors, *ACM CCS 15: 22nd Conference on Computer and Communications Security*, pages 644–655, Denver, CO, USA, October 12–16, 2015. ACM Press.

[33] Muhammad Naveed, Manoj Prabhakaran, and Carl A. Gunter. Dynamic searchable encryption via blind storage. In *2014 IEEE Symposium on Security and Privacy*, pages 639–654, Berkeley, CA, USA, May 18–21, 2014. IEEE Computer Society Press.

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

[35] Dong Jin Park, Kihyun Kim, and Pil Joong Lee. Public key encryption with conjunctive field keyword search. In Chae Hoon Lim and Moti Yung, editors, *WISA 04: 5th International Workshop on Information Security Applications*, volume 3325 of *Lecture Notes in Computer Science*, pages 73–86, Jeju Island, Korea, August 23–25, 2004. Springer, Heidelberg, Germany.

[36] David Pouliot and Charles V. Wright. The shadow nemesis: Inference attacks on efficiently deployable, efficiently searchable encryption. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *ACM CCS 16: 23rd Conference on Computer and Communications Security*, pages 1341–1352, Vienna, Austria, October 24–28, 2016. ACM Press.

[37] Elaine Shi, John Bethencourt, Hubert T.-H. Chan, Dawn Xiaodong Song, and Adrian Perrig. Multi-dimensional range query over encrypted data. In *2007 IEEE Symposium on Security and Privacy*, pages 350–364, Oakland, CA, USA, May 20–23, 2007. IEEE Computer Society Press.

[38] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *2000 IEEE Symposium on Security and Privacy*, pages 44–55, Oakland, CA, USA, May 2000. IEEE Computer Society Press.

[39] Emil Stefanov, Charalampos Papamanthou, and Elaine Shi. Practical dynamic searchable encryption with small leakage. In *ISOC Network and Distributed System Security Symposium – NDSS 2014*, San Diego, CA, USA, February 23–26, 2014. The Internet Society.

[40] J. Wang, X. Chen, X. Huang, I. You, and Y. Xiang. Verifiable auditing for outsourced database in cloud computing. *IEEE Transactions on Computers*, 64(11):3293–3303, Nov 2015.

[41] Peishun Wang, Huaxiong Wang, and Josef Pieprzyk. Keyword field-free conjunctive keyword searches on encrypted data and extension for dynamic groups. In Matthew K. Franklin, Lucas Chi Kwong Hui, and Duncan S. Wong, editors, *CANS 08: 7th International Conference on Cryptology and Network Security*, volume 5339 of *Lecture Notes in Computer Science*, pages 178–195, Hong-Kong, China, December 2–4, 2008. Springer, Heidelberg, Germany.

[42] Peng Xu and Hai Jin. Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack. Cryptology ePrint Archive, Report 2010/626, 2010. `http://eprint.iacr.org/2010/626`.

[43] Qingji Zheng, Shouhuai Xu, and Giuseppe Ateniese. VABKS: Verifiable attribute-based keyword search over outsourced encrypted data. Cryptology ePrint Archive, Report 2013/462, 2013. `http://eprint.iacr.org/2013/462`.

[44] Hong Zhu, Zhuolin Mei, Bing Wu, Hongbo Li, and Zongmin Cui. Fuzzy keyword search and access control over ciphertexts in cloud computing. In Josef Pieprzyk and Suriadi Suriadi, editors, *ACISP 17: 22nd Australasian Conference on Information Security and Privacy, Part I*, volume 10342 of *Lecture Notes in Computer Science*, pages 248–265, Auckland, New Zealand, July 3–5, 2017. Springer, Heidelberg, Germany.