与上述方案相比,VOKCRSII 方案不仅支持多关键字搜索结果排序与可验证的功能,还支持插入混淆关键字,使方案安全性更高.

Table 1 Function comparison

表 1 功能比较

功能	文献		
	文献[9]	文献[8]	VOKCRSII
结果排序	√	√	√
可验证	√	√	√
混淆关键字	–	–	√

4.2 安全性

实验利用公式(9)检测文献[8]、文献[9]和 VOKCRSII 方案的隐私保护水平.

$$H(D)=-\sum_{i=1}^{m}p(D_i)\log_2 p(D_i) \tag{9}$$

其中,$0<p(D_i)<1$, $\sum_{i=1}^{m}p(D_i)=1$.

$H(D)$越大,隐私泄露可能性就越小,在没有外部条件影响时,该值是一个确定的值[17].

如图 5 所示是文献[8]、文献[9]与 x=2 时的 VOKCRSII 隐私保护度对比.VOKCRSII 赋予用户可验证的权利,且在查询陷门中引进混淆关键字,防止云服务器恶意攻击搜索频率高的数据,或者删除搜索频率低的数据,因此,x=2 时,VOKCRSII 方案的隐私保护度高于文献[9]和文献[8]提出的方案,安全性更高.

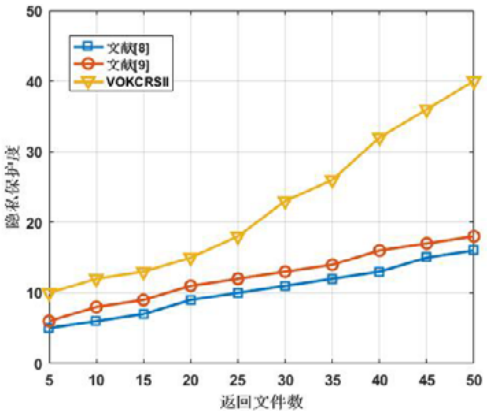


Fig.5 Comparison chart of privacy protection

图 5 隐私保护度比较图

4.3 检索效率

授权用户在进行检索时,总希望快速地得到检索结果[18],本文分别对方案的生成陷门时间、查询时间和验证时间进行实验.由第 4.2 节可知,x=2 时,VOKCRSII 方案的安全性已经高于文献[9]和文献[8]的方案.随着插入混淆关键字个数的增长,安全性会增加,但会带来更多的计算开销.因此选择引入 2 个混淆关键字的 VOKCRSII 方案做对比实验.

1) 数据缓存区

VOKCRSII 比文献[9]方案多消耗的时间主要体现在云服务器将查询数据映射到数据缓存区过滤掉包含混淆关键字和利用 Paillier 解密恢复数据两个方面.表 2 的第 2 行是插入不同数量混淆关键字时映射和过滤消耗的时间,当 x=2 时,时间为 0.160s.表 2 的第 3 行是随着插入混淆关键字数量增加利用 Paillier 解密恢复数据的时间.当 x=2 时,时间为 0.031s.

Table 2 Time of mapping filter and decryption data

表 2 映射过滤和解密时间

混淆关键字个数	1	2	3	4	5	6	7	8	9
映射过滤时间(s)	0.081	0.160	0.252	0.325	0.401	0.474	0.582	0.701	0.823
Paillier解密时间(s)	0.023	0.031	0.034	0.042	0.045	0.043	0.047	0.051	0.051

2) 查询效率

查询过程可以分为陷门生成、查询和验证 3 个部分.

生成陷门时,如表 3 所示,文献[8]的方案用到了大量的内积运算,复杂度是 $O(n^2)$,与关键词集大小有关.文献[9]的方案中,陷门只是由 3 个 PRF 产生的 3 个伪随机位序列组成.构造陷门的复杂度是 $O(\lambda)$.VOKCRSII 虽然在陷门中加入了混淆关键字,但构造陷门的复杂度仍是 $O(\lambda)$.如图 6,VOKCRSII 和文献[9]的构造陷门时间只与随机种子λ相关,而与 n 无关,随着关键词数量的增加时间几乎不变.

Table 3 Comparison of time complexity

表 3 时间复杂度比较

Scheme	Trapdoor	Search	Verify
文献[8]	$O(n^2)$	$O(nm)$	$O(nm)+O(m\log m)$
文献[9]	$O(\lambda)$	$O(\#w)$	$O(\#w+\Sigma\#C_{top\text{-}K})+O(\Sigma\#w)$
VOKCRSII	$O(\lambda)$	$O(xK)$	$O(\#w+\lambda)+O(\#w)$

查询时,文献[8]涉及搜索陷门和每个文档子索引的内积,见表 3,查询时间的复杂度为 $O(nm)$.由于倒排索引搜索的时间成本与包含 w 的文档的数量成线性关系,文献[10]查询时间复杂度为 $O(\#w)$,VOKCRSII 由于要搜索 x 混淆关键字的文件以及将数据映射到数据缓存区,因此查询时间的复杂度为 $O(xK)$.如图 7 所示,由于文献[8]的方案检索时间随着文档数量的增加而增加,时间最长.而文献[9]与 VOKCRSII 只与包含搜索关键字的文件数量相关,相较文献[8]的方案查询时间增长缓慢,其中,VOKCRSII 引入了混淆关键字来提高检索的安全性,需要过滤混淆关键字,VOKCRSII 比文献[9]的方案的所用时间长,但相差不是很多.

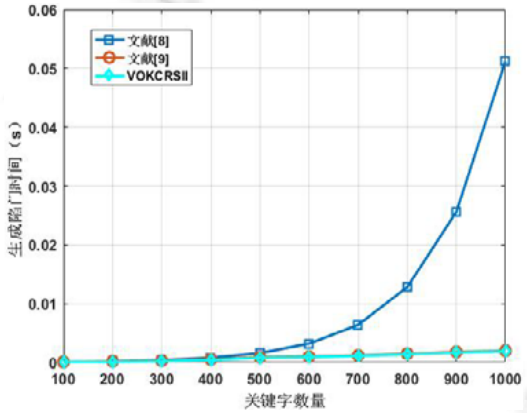


Fig.6 Time of generate trapdoor

图 6 陷门生成时间

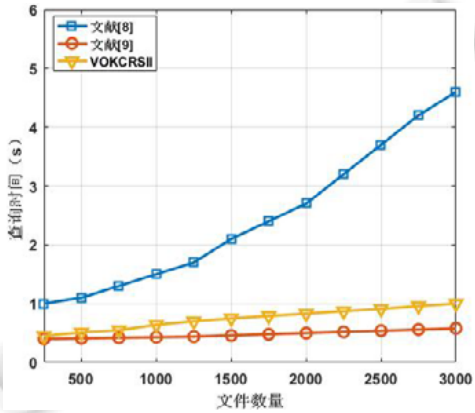
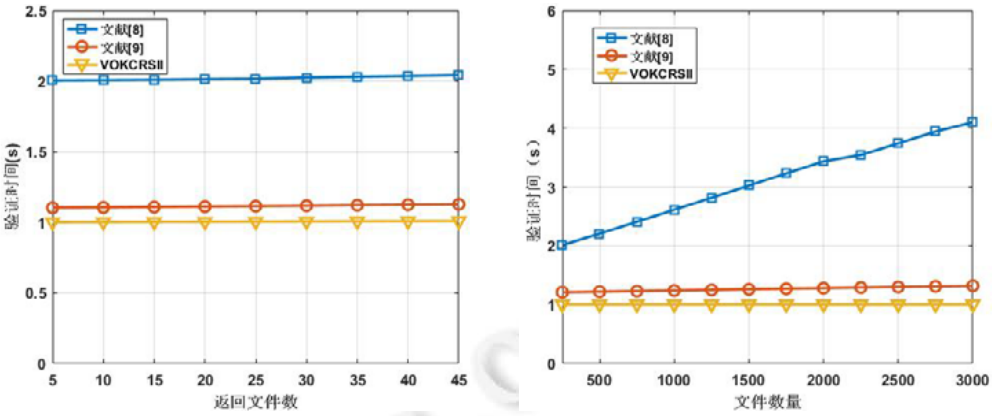


Fig.7 Search time

图 7 查询时间

验证时间包括在云服务器端生成标签的时间与用户验证的时间两部分.如表 3 所示,由于文献[8]要计算文档之间的向量积,在客户端验证搜索结果的复杂度为 $O(nm)$;云服务器端通过 hash 验证树生成标签,时间复杂度为 $O(m\log m)$.文献[9]在云服务器链签名技术生成标签,时间复杂度为 $O(\Sigma\#w)$;收到来自云服务器的返回结果和标签后,客户端利用 MAC 将查询关键字和返回的 top-K 文档的连接作为输入进行验证,复杂度为 $O(\#w+\Sigma\#C_{top\text{-}K})$,其中,$\#w$ 表示查询关键字的长度,$\#C_{top\text{-}K}$ 表示返回的 top-K 文档的总长度.VOKCRSII 在云服务器生成标签时间复杂度为 $O(\#w)$;在客户端数据用户先利用双线性映射的性质确定返回的结果是否是包含关键字 w_s

的文件,再验证来确定返回的结果是否正确,复杂度为 $O(\#w+\lambda)$.如图 8(a)所示,由于文献[8]利用 MAC 来验证,验证时间最长.文献[9]验证的复杂度与 top-K 文档的总长度相关,随着用户要求返回文档数量的增加,检索时间增长.由于 VOKCRSII 引入混淆关键字,映射过滤消耗的时间要随之增加,但 VOKCRSII 验证时不涉及对返回密文的计算,验证时间最短.如图 8(b)所示,当 Top-K=20 时,随着文件集数量变化,文献[9]验证消耗时间呈线性增长.而文献[9]和 VOKCRSII 与包含查询关键字的文档数量有关,验证时间增长缓慢.



(a) 不同的返回文档数量对应的验证时间 (b) 不同的文档数量对应的验证时间

Fig.8 Verification time

图 8 验证时间

5 总 结

本文提出了一个安全有效的关键字密文检索方案 VOKCRSII.该方案通过混淆关键字隐藏搜索频率,并利用双线性映射生成标签验证搜索结果,提高方案的安全性.同时,对其正确性、安全性和可靠性这 3 个方面进行了验证.通过分析验证,VOKCRSII 满足自适应性选择关键字攻击安全.此外,利用加密标志位区分混淆关键字和真正要检索的关键字,生成陷门上传至云服务器,但根据陷门得到的搜索结果有包含混淆关键字的文件,利用 Paillier 加密算法生成数据缓存区过滤掉多余文件,以减少通信开销.通过建立密文检索实验平台,验证 VOKCRSII 在保证检索效率的同时,提高了密文检索的安全性.但 VOKCRSII 只支持加密文档集的查询,将可搜索加密技术扩展到关系型数据库,是未来要做的工作.

References:

[1] Li JW, Jia CF, Liu ZL, Li J, Li M. Survey on the searchable encryption. Ruan Jian Xue Bao/Journal of Software, 2015, 26(1):109−128 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4700.htm [doi: 10.13328/j.cnki.jos.004700]

[2] Song XD, Wagner D, Perrig A. Practical techniques for searches on encrypted data. In: Proc. of the IEEE Symp. on Security and Privacy. IEEE Press, 2000. 44−55.

[3] Curtmola R, Garay J, Kamara S, Ostrovsky R. Searchable symmetric encryption: Improved definitions and efficient constructions. In: Proc. of the 13th ACM Conf. on Computer and Communications Security. New York: ACM Press, 2006. 79−88.

[4] Ibrahim A, Jin H, Yassin AA, *et al*. Secure rank-ordered search of multi-keyword trapdoor over encrypted cloud data. In: Proc. of the IEEE Asia-Pacific Services Computing Conf. IEEE Computer Society, 2012. 263−270.

[5] Chen XF, Huang XY, Li J, *et al*. New algorithms for secure outsourcing of large-scale systems of linear equations. IEEE Trans. on Information Forensics & Security, 2014,10(1):69−78.

[6] Sun W, Wang B, Cao N, *et al*. Privacy-Preserving multi-keyword text search in the cloud supporting similarity-based ranking. In: Proc. of the ACM Sigsac Symp. on Information, Computer and Communications Security. ACM Press, 2013. 71−82.

[7] Chen C, Zhu X, Shen P, *et al*. An efficient privacy-preserving ranked keywordssearch method. IEEE Trans. on Parallel & Distributed Systems, 2016,27(4):951−963.

[8] Jiang X, Yu J, Yan J, *et al*. Enabling efficient and verifiable multi-keyword ranked search over encrypted cloud data. Information Sciences, 2017,403(3):22−41.

[9] Liu Q, Nie X, Liu X, *et al*. Verifiable ranked search over dynamic encrypted data in cloud computing. In: Proc. of the Int'l Symp. on Quality of Service. IEEE, 2017. 1−6.

[10] Zhang W, Lin Y, Gu Q. Catch you if you misbehave: ranked keyword search results verification in cloud computing. IEEE Trans. on Cloud Computing, 2018,1(6):74−86.

[11] Wan Z, Deng RH. VPSearch: Achieving verifiability for privacy-preserving multi-keywordsearch over encrypted cloud data. IEEE Trans. on Dependable & Secure Computing, 2018,15(6):1083−1095.

[12] Zhang R, Xue R, Yu T, *et al*. PVSAE: A public verifiable searchable encryption service framework for outsourced encrypted data. In: Proc. of the IEEE Int'l Conf. on Web Services. IEEE, 2016. 428−435.

[13] Qiu S. Research on privacy-preserving keyword search and set operations over encrypted data [Ph.D. Thesis]. Beijing: Beijing Jiaotong University, 2017 (in Chinese with English abstract).

[14] Wu ZQ, Li KL, Zheng H. Efficient and scalable architecture forsearchable symmetric encryption. Journal on Communications, 2017,38(8):79−93 (in Chinese with English abstract).

[15] Wang SP, Liu LJ, Zhang YL. Verifiable dictionary-based searchable encryption scheme. Ruan Jian Xue Bao/Journal of Software, 2016,27(05):1301−1308 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4912.htm [doi: 10.13328/j.cnki.jos.004912]

[16] Du MX, Wang Q, He MQ, *et al*. Privacy-Preserving indexing and query processing for secure dynamic cloud storage. IEEE Trans. on Information Forensics and Security, 2018,13(9):2320−2332.

[17] Peng CG, Ding HF, Zhu YJ, Tian YL, Fu ZF. Information entropy models and privacy metrics methods for privacy protection. Ruan Jian Xue Bao/Journal of Software, 2016,27(8):1891−190 (in Chinese with English abstract). http://www.jos. org.cn/1000-9825/5096.htm [doi: 10.13328/j.cnki.jos.005096]

[18] Dong XL, Zhou J, Cao ZF. Research advances on secure searchable encryption. Journal of Computer Research and Development, 2017,54(10):2107−2120 (in Chinese with English abstract).

附中文参考文献:

[1] 李经纬,贾春福,刘哲理,李进,李敏.可搜索加密技术研究综述.软件学报,2015,26(1):109−128. http://www.jos.org.cn/1000-9825/4700.htm [doi: 10.13328/j.cnki.jos.004700]

[13] 邱硕.面向隐私保护的密文数据检索与集合操作的关键技术研究[博士学位论文].北京:北京交通大学,2017.

[14] 吴志强,李肯立,郑蕙.高效可扩展的对称密文检索架构.通信学报,2017,38(8):79−93.

[15] 王尚平,刘利军,张亚玲.可验证的基于词典的可搜索加密方案.软件学报,2016,27(5):1301−1308. http://www.jos.org.cn/1000-9825/4912.htm [doi: 10.13328/j.cnki.jos.004912]

[17] 彭长根,丁红发,朱义杰,田有亮,符祖峰.隐私保护的信息熵模型及其度量方法.软件学报,2016,27(8):1891−1903. http://www.jos.org.cn/1000-9825/5096.htm [doi: 10.13328/j.cnki.jos.005096]

[18] 董晓蕾,周俊,曹珍富.可搜索加密研究进展.计算机研究与发展,2017,54(10):2107−2120.

杜瑞忠(1975－),男,河北献县人,博士,教授,CCF 专业会员,主要研究领域为可信计算,信息安全.

田俊峰(1975－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为分布计算,可信计算,信息安全.

李明月(1993－),女,硕士生,主要研究领域为可信计算,信息安全.

吴万青(1981－),男,博士,讲师,主要研究领域为信息安全,密码学.