3.2 密文数据库的关系运算

经过 OPE 的数据库并不能实现所有密文域上的数据库关系运算,例如,SUM,AVG 等运算需要解密后才能实现.但由于保持顺序的关系,仍可以在不解密情况下进行等式和范围查询(如 MAX,MIN,COUNT,GROUPBY,ORDERBY)等关系运算.图 5 显示了在加密数据库下用户查询处理的过程,查询 SQL 语句经过翻译层时,对数值型数据和字符型数据分别进行 OPES 和 OPES+,数据库系统返回符合请求的加密数据,在翻译层中进行解密并返回给用户.

Fig.5 Process of query processing in encrypted database

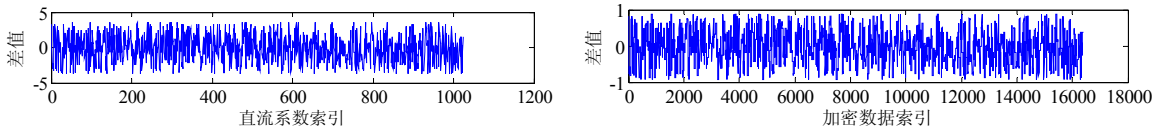
图 5 加密数据库查询处理

4 实验结果及分析

本文算法使用版本为 1.8.0 的 JVM 运行在配置为 3.50GHz Intel i5 处理器,4GB 运行内存的 Windows 7 系统上,通过 JDBC 方式连接到 MySQL 5.6 数据库,实现数据的增、删、改、查操作.首先建立一个数据库,并建立一张含有几个属性列的表,实验选取图像大小为 128×128 的 Lena 灰度图像的像素值作为实验数据.

4.1 水印生成及嵌入

将 128×128 个像素值数据写入数据库某个数值型属性列,进行保序加密得到加密数据.信息隐藏者对加密数据进行分组,每组 16 个数据,共 1 024 组,对每组进行 DCT 变换后获取其后 8 个交流系数,得到 8 个由 1 024 个交流系数组成的交流系数序列,每个序列进行哈希分别得到 128 比特的二值信息,然后对这些信息进行合并,得到 1 024 比特的水印信息.利用 QIM 算法将水印信息分别嵌入到 1 024 个直流系数中,根据公式(23)可知,量化步长Δ取值范围为(0,16/3),这里取Δ=5,然后将含水印的密文更新到数据库.图 6(a)显示了将 1 024 比特的水印信息分别嵌入到 1 024 个直流系数前后的差值,范围在(−4,4)之间;图 6(b)显示加密数据库嵌入水印前后的差值,水印嵌入失真刚好控制在(−1,1)范围内,显示了前面理论分析的正确性.



(a) 直流系数嵌水印前后的差值 (b) 加密数据库数据嵌水印前后的差值

Fig.6 Watermark distortion of DC coefficients and encrypted data

图 6 直流系数和加密数据的水印失真

4.2 完整性认证及篡改对数据库恢复影响

在接收端,接收者可对加密数据库进行完整性认证,也可根据密钥直接对数据库进行解密,得到原明文数据库.为了测试篡改对数据库恢复的影响,表 2 显示了随机选取 5 个含水印的密文值进行了不同程度的篡改,实验结果如图 7 所示.

Table 2 Tampering partial data

表 2 部分数据篡改

索引号	含水印密文值	篡改程度	篡改后含水印密文值
578	3 946.187 5	+0.1	3 946.287 5
1 575	3 912.500 0	−5	3 907.500 0
2 291	835.062 5	+50	885.062 5
10 119	697.562 5	−200	497.562 5
13 906	5 913.562 5	+350	6 263.562 5

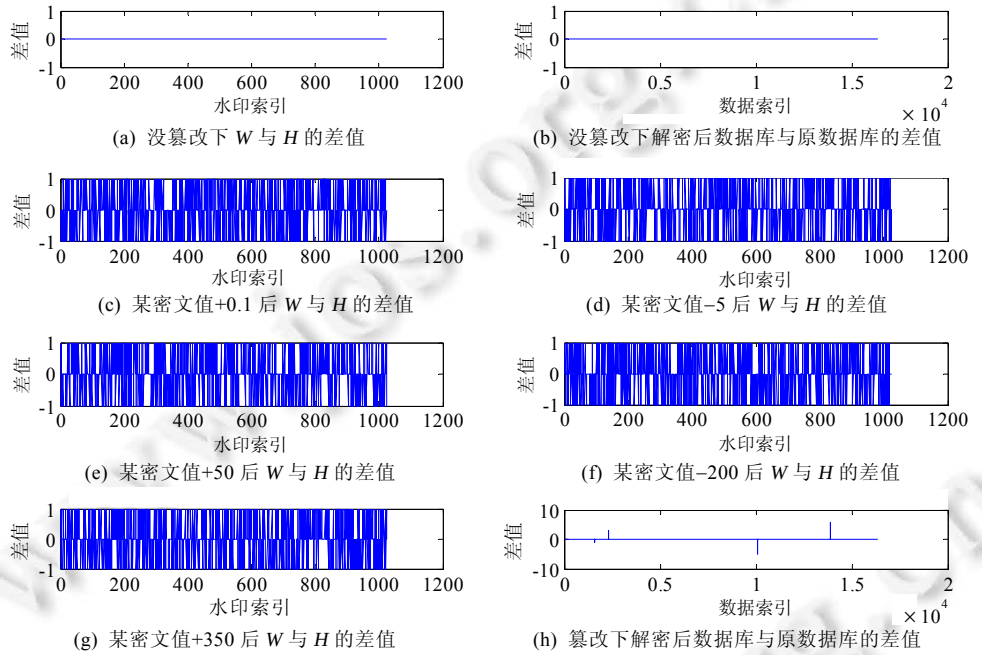


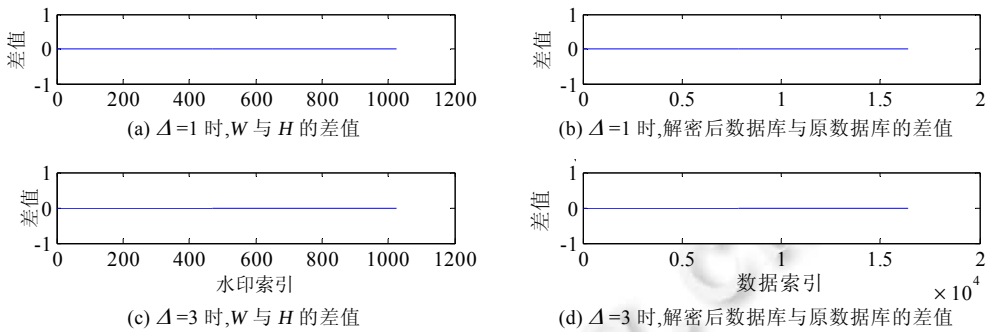
Fig.7 Tamper detection experiment: by comparing the watermark information in DC coefficients and the hash value generated by AC coefficients to determine whether database has been tampered with

图 7 篡改检测实验:通过比对直流系数中的水印信息(*W*)和交流系数的哈希值(*H*)来判断是否被篡改

图 7(a)显示在没有篡改的情况下,含水印加密数据库完整性认证情况,差值均为 0,可以看出,提取出来的水印信息与生成的对比信息完全相同.图 7(b)表示在不篡改情况下,含水印加密数据库的明文恢复情况,差值也均为 0,即,解密后的数据库与原数据库完全相同,说明对加密数据进行水印嵌入不影响明文数据的恢复.图 7(c)~图 7(g)分别表示对表 2 中的一个含水印密文进行不同程度的篡改操作后的认证情况.可以看出:即使是很小幅度的篡改,直流系数中提取出来的水印信息与交流系数生成的哈希也是完全不同的,这充分显示了本文加密域认证水印算法的有效性.图 7(h)表示在表 2 所有密文进行篡改的情况下,对含水印加密数据库进行解密后得到的明文情况,可以看出有 4 处错误(除了+0.1 的篡改),这说明了认证水印对于数据库完整性保护的必要性.通过在加密域引入认证水印,可以保护敏感数据的安全,防止非法篡改.

4.3 不同步长选取对认证水印算法及明文恢复的影响

公式(23)表明,量化步长Δ允许在一定范围内取值.下面通过实验分析不同步长的选取对认证水印算法及明文数据库恢复的影响.随机选取了 2 个不同的量化步长进行实验,分别为Δ=1 和Δ=3.实验结果如图 8 所示.图 8(a)~图 8(d)分别表示了选取不同量化步长下,含水印加密数据库完整性认证情况和明文恢复情况,差值均为 0.结果表明:在公式(23)的约束下,量化步长Δ的选取不会影响认证水印算法以及明文数据库的恢复.



(a) Δ=1 时,W 与 H 的差值　(b) Δ=1 时,解密后数据库与原数据库的差值

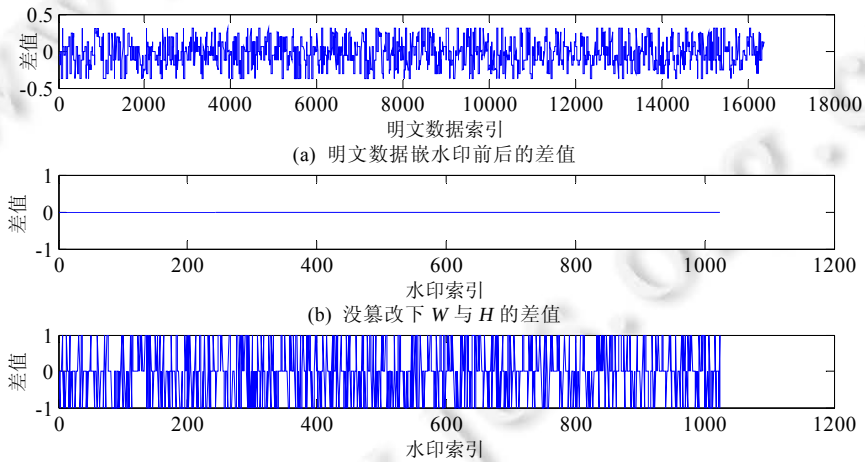
(c) Δ=3 时,W 与 H 的差值　(d) Δ=3 时,解密后数据库与原数据库的差值

Fig.8　Influence of different Δ on authentication watermark algorithm and database recovery

图 8　不同步长选取对认证水印算法及明文恢复的影响

4.4　认证水印算法对明文数据库的可用性测试

下面测试本文认证水印算法对于明文数据库的可用性,明文数据同样是 128×128 的像素值.为了保持明文顺序不变,可篡改范围设为(–0.5,0.5),由公式(23)计算量化步长为Δ=2.测试结果如图 9 所示.图 9(a)显示明文数据库嵌入水印前后的差值,水印嵌入失真刚好控制在(–0.5,0.5)范围内,即,水印的嵌入没有改变明文数据的顺序,印证了公式(23)的正确性;图 9(b)显示在没有篡改的情况下,含水印明文数据库完整性认证情况,可以看出,提取出来的水印信息与生成的对比信息完全相同;图 9(c)表示随机选取一个含水印明文,并进行+5 篡改操作后的认证情况,可以看出,明文数据库下本文认证算法仍然有效.



(a) 明文数据嵌水印前后的差值

(b) 没篡改下 W 与 H 的差值

(c) 随机取一个含水印明文进行+5 篡改后,W 与 H 的差值

Fig.9　Usability testing of authentication watermarking algorithm in original database

图 9　认证水印算法对明文数据库的可用性测试

从实际意义上来说,在明文数据库上直接嵌入水印并不明智,因为明文数据已经遭到了破坏,尤其对于准确性要求较高的数据库.通过保序概率加密和其引入的冗余,本文达到了保护隐私、嵌入认证信息和进行关系运算等目的.

4.5　时间开销

下面分析在资源受限的用户端下,本文的保序概率加密算法和认证水印算法的时间开销(计算时间开销和存储时间开销).为了测试方便,本文实验的数据库表只含有两个属性列(自增的索引列和待加密数据列),每一行数据代表一个元组.

图 10 显示了文献[10,12]算法(本文所用算法)在不同元组数下,加/解密所需的时间开销.可以看出:随着处理元组数量越多,加密和解密所需的时间开销也随之增加.可以看出:本文所用的保序概率加密算法在时间开销上比文献[10]要大一些,但在安全性上比文献[10]要好得多.

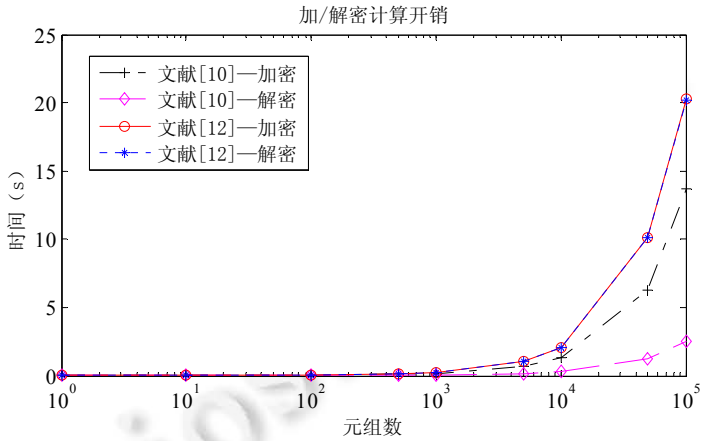


Fig.10 Time required for encrypting/decrypting different number of tuples

图 10 加/解密不同元组数所需时间

图 11 显示了一次性插入或读取不同数量的元组所需要的时间开销,可以看出:随着插入和读取的元组数量增加,所需要的时间开销也随之增加,其中:插入元组时时间开销增加非常快,而读取元组时则很缓慢.这是因为插入操作会改动原有数据库,而查询仅仅是返回所需数据.

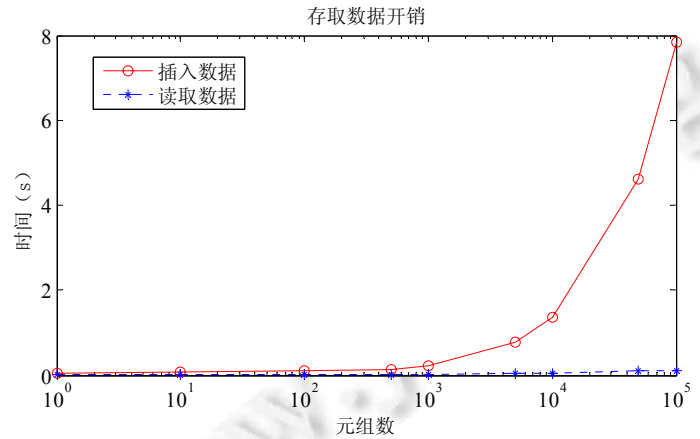


Fig.11 Time needed to one-time insert/read different number of tuples

图 11 一次性插入或读取不同元组数所需的时间

图 12 显示了在不同元组数下,本文认证水印算法的水印嵌入过程和水印提取过程所需要的时间开销.水印嵌入的时间由水印嵌入算法运行时间和含水印数据更新到数据库的时间组成.结合图 11 可知,水印嵌入的时间主要取决于后者所用的时间.水印提取不涉及数据库更改,其时间开销由水印提取算法运行时间和数据读取时间组成,随着元组数的增加,时间开销增加缓慢.

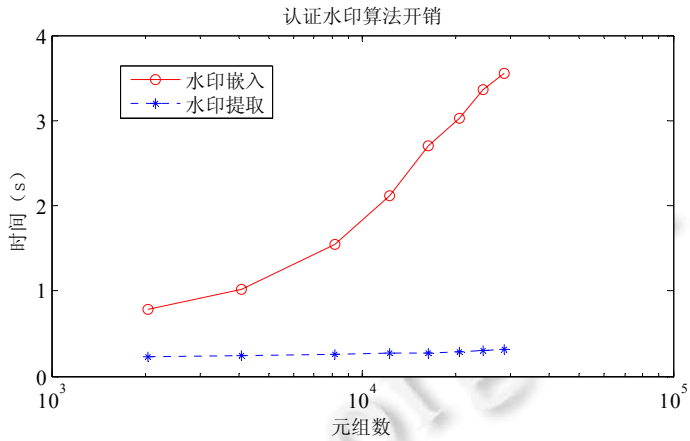


Fig.12 Time overhead of authentication watermarking algorithm in different number of tuples

图 12 不同元组数的认证水印算法时间开销

5 结 论

本文实现了一种新的保序加密域数据库认证水印算法.数据库所有者首先对数据库某一属性列的敏感数据进行加密,以保护其内容隐私;信息隐藏者对加密数据进行水印嵌入处理,防止数据被非法篡改,水印嵌入不影响用户对数据的查询和使用;接收者可通过验证数据库的完整性来确保是否被篡改,对于拥有加密密钥的接收者,可直接对加密数据库进行解密,得到原数据库.本文算法解决了以下几点问题并获得了很好的效果.

1) 通常,明文数据库中数据冗余非常小,进行水印保护时难以找到用于嵌入的冗余空间.经过保序加密的数据库数据不仅保护了内容隐私,而且所有数据都可以进行水印处理,很好地解决了数据库的保护问题;
2) 水印嵌入过程很好地结合了保序加密的特点,对加密数据的水印嵌入相对于明文数据是无损的,对含水印的密文数据库直接解密可得到原数据库.好处是:水印对加密数据提供了认证保护,但不会影响数据库的使用;
3) 水印算法很好地利用了 DCT 的全局和正交特性,将交流系数的哈希值嵌入直流系数,达到了对数据库的完整性认证,能够识别不同程度的篡改,对加密数据的完整性提供了很好的保护.

本文算法适合于网络空间安全大背景下的云数据库服务,很好地结合了加密和信息隐藏技术,为数据内容的隐私保护和数据安全提供了一种有效的潜在技术手段.

References:

[1] Zhang XP, Long J, Wang ZC, Cheng H. Lossless and reversible data hiding in encrypted images with public key cryptography. IEEE Trans. on Circuits and Systems for Video Technology, 2016,26(9):1622−1631. [doi: 10.1109/TCSVT.2015.2433194]

[2] Xiang SJ, Luo XR. Reversible data hiding in encrypted image based on homomorphic public key cryptosystem. Ruan Jian Xue Bao/ Journal of Software, 2016,27(6):1592−1601 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5007.htm [doi: 10.13328/j.cnki.jos.005007]

[3] Huang LS, Tian MM, Huang H. Preserving privacy in big data: A survey from the cryptographic perspective. Ruan Jian Xue Bao/ Journal of Software, 2015,26(4):945−959 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4794.htm [doi: 10.13328/j.cnki.jos.004794]

[4] Marisa P, Andry A, Budi R. Big-Data security management issue. In: Proc. of the 2nd Int'l Conf. on Information and Communication Technology (ICoICT). Bandung, 2014. 59−63. [doi: 10.1109/ICoICT.2014.6914040]

[5] Tian XX, Wang XL, Gao M, Zhou AY. Database as a service—Security and privacy preserving. Ruan Jian Xue Bao/Journal of Software, 2010,21(5):991−1006 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3746.htm [doi: 10.3724/SP.J.1001.2010.03746]

[6] Mario G. New challenges in teaching database security. In: Proc. of the 3rd Annual Conf. on Information Security Curriculum Development. Kennesaw, 2006. 64−67. [doi: 10.1145/1231047.1231060]

[7] Murray MC. Database security: What students need to know. Journal of Information Technology Education, 2010,9:61−77.

[8] Du L, Cao XC, Zhang W, Zhang XP, Liu N, Wei JG. Semi-Fragile watermarking for image authentication based on compressive sensing. Science China Information Sciences, 2016,59(5):1−3. [doi: 10.1007/s11432-016-5542-8]

[9] Schmitz R, Li SJ, Grecos C, Zhang XP. Content-Fragile commutative watermarking-encryption based on pixel entropy. Springer Int'l Publishing, 2015,9386:474−485. [doi: 10.1007/978-3-319-25903-1_41]

[10] Agrawal A, Kiernan J, Srikant R, Xu YR. Order preserving encryption for numeric data. In: Proc. of the 2004 ACM SIGMOD Int'l Conf. on Management of Data. New York, 2004. 563−574. [doi: 10.1145/1007568.1007632]

[11] Ma YZ, Meng XF. Research on indexing for cloud data management. Ruan Jian Xue Bao/Journal of Software, 2015,26(1):145−166 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4688.htm [doi: 10.13328/j.cnki.jos.004688]

[12] Boldyreva A, Chenette N, Lee Y, O'Neill A. Order-Preserving symmetric encryption. In: Proc. of the 28th Annual Int'l Conf. on Advances in Cryptology. Berlin, 2009. 224−241. [doi: 10.1007/978-3-642-01001-9_13]

[13] Wang C, Cao N, Li J, Ren K, Lou W. Secure ranked keyword search over encrypted cloud data. In: Proc. of the 30th IEEE Int'l Conf. on Distributed Computing Systems. Genova, 2010. 253−262. [doi: 10.1109/ICDCS.2010.34]

[14] Tang Q. Privacy preserving mapping schemes supporting comparison. In: Proc. of the 2010 ACM Workshop on Cloud Computing Security. New York, 2010. 53−58. [doi: 10.1145/1866835.1866846]

[15] Boldyreva A, Chenette N, O'Neill A. Order-Preserving encryption revisited: Improved security analysis and alternative solutions. In: Proc. of the 31st Annual Int'l Conf. on Advances in Cryptology. Santa Barbara, 2011. 578−595. [doi: 10.1007/978-3-642-22792-9_33]

[16] Wang C, Cao N, Ren K, Lou WJ. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans. on Parallel and Distributed Systems, 2012,23(8):1467−1479. [doi: 10.1109/TPDS.2011.282]

[17] Li K, Zhang WM, Yang C, Yu NH. Security analysis on one-to-many order preserving encryption-based cloud data search. IEEE Trans. on Information Forensics and Security, 2015,10(9):1918−1926. [doi: 10.1109/TIFS.2015.2435697]

[18] Popa RA, Li FH, Zeldovich N. An ideal-security protocol for order-preserving encoding. In: Proc. of the 2013 IEEE Symp. on Security and Privacy. Berkeley, 2013. 463−477. [doi: 10.1109/SP.2013.38]

[19] Agrawal R, Kiernan J. Watermarking relational databases. In: Proc. of the 28th Int'l Conf. on Very Large Databases. Hong Kong, 2002. 155−166. [doi: 10.1016/B978-155860869-6/50022-6]

[20] Sion R, Atallah M, Prabhakar S. Rights protection for relational data. IEEE Trans. on Knowledge and Data Engineering, 2004, 16(12):1509−1525. [doi: 10.1109/TKDE.2004.94]

[21] Zhou F, Zhao HX. Relational database watermarking algorithm based on chaos and DCT. Application Research of Computers, 2012, 29(2):786−788 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2012.02.104]

[22] Huang C, Zhu YL. Fast algorithm for arbitrary length discrete cosine transform. In: Proc. of the 5th Int'l Conf. on Natural Computation. Tianjin, 2009. 390−393. [doi: 10.1109/ICNC.2009.640]

[23] Chen B, Wornell GW. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Trans. on Information Theory, 2001,47(4):1423−1443. [doi: 10.1109/18.923725]

[24] Li YX, Liu GH. Order preserving encryption method for character data in relational databases. Radio Engineering, 2006,36(4):1−3 (in Chinese with English abstract).

附中文参考文献:

[2] 项世军,罗欣荣.基于同态公钥加密系统的图像可逆信息隐藏算法.软件学报,2016,27(6):1592−1601. http://www.jos.org.cn/1000-9825/5007.htm [doi: 10.13328/j.cnki.jos.005007]

[3] 黄刘生,田苗苗,黄河.大数据隐私保护密码技术研究综述.软件学报,2015,26(4):945−959. http://www.jos.org.cn/1000-9825/4794.htm [doi: 10.13328/j.cnki.jos.004794]

[5] 田秀霞,王晓玲,高明,周傲英.数据库服务——安全与隐私保护.软件学报,2010,21(5):991−1006. http://www.jos.org.cn/1000-9825/3746.htm [doi: 10.3724/S P.J.1001.2010.03746]

[11] 马友忠,孟小峰.云数据管理索引技术研究.软件学报,2015,26(1):145−166. http://www.jos.org.cn/1000-9825/4688.htm [doi: 10.13328/j.cnki.jos.004688]

[21] 周飞,赵怀勋.基于混沌的 DCT 域关系数据库水印算法.计算机应用研究,2012,29(2):786−788. [doi: 10.3969/j.issn.1001-3695.2012.02.104]

[24] 李亚秀,刘国华.关系数据库中字符数据的保序加密方法.无线电工程,2006,36(4):1−3.

项世军(1974－),男,贵州普定人,博士,教授,CCF 高级会员,主要研究领域为信息隐藏,加密域信号处理.

何嘉勇(1992－),男,硕士,主要研究领域为多媒体信息安全,加密域信号处理.