文章编号：2095-4980(2020)05-0929-05

基于改进 RSA 算法的隐私数据集同态加密方法

鲍海燕，芦彩林

(晋中学院 信息技术与工程学院，山西 晋中 030619)

摘 要：为提高对数据集的加密效果和保障数据集的安全性，在分析传统非对称密钥(RSA)算法运行原理及其参数选择、素数判定条件的基础上，提出了改进RSA算法。为进一步提升加密速度，引入数据加密(DES)算法。首先利用DES算法实现明文数据集的加密，并针对密钥进行RSA加密；在此基础上，在明文和密文空间中，利用加法同态过程对密文进行计算，并通过对结果的解密操作得到相应的明文计算结果。实验结果表明，与基于传统RSA算法或DES算法的加密方法相比，该方法的加密效率和抵御攻击成功率更高，加密过程耗时5~6 s，抵御攻击成功率保持在95%上下，说明该方法能够有效保护隐私数据集的安全。

关键词：改进非对称密钥(RSA)算法；数据加密(DES)算法；混合算法；隐私数据集；同态加密

中图分类号：TN918.4 **文献标志码：**A **doi：**10.11805/TKYDA2020072

Homomorphic encryption of privacy data set based on improved RSA algorithm

BAO Haiyan，LU Cailin

(School of Information Technology and Engineering，Jinzhong University，Jinzhong Shanxi 030619，China)

Abstract：In order to improve the encryption effect and ensure the security of the data set, this study designs a privacy data set homomorphic encryption method based on the improved Rivest Shamir Adleman(RSA) algorithm. Based on the analysis of the operating principle of traditional RSA algorithm, the parameter selection and the prime number judgment condition, the improved RSA algorithm is proposed. In order to further improve the encryption speed, Data Encryption Standard(DES) algorithm is introduced. Firstly, DES algorithm is utilized to encrypt the plaintext data set, and RSA encryption is carried out for the key. On this basis, in the plaintext and ciphertext spaces, the addition homomorphism process is adopted to calculate the ciphertext, and the corresponding plaintext calculation result is obtained by decrypting the result. Experimental results show that, compared with the encryption method based on traditional RSA algorithm or DES algorithm, this method has higher encryption efficiency and higher success rate of resisting attacks. The encryption process takes time between 5 and 6 s, and the success rate of resisting attacks is kept around 95%, indicating that this method can effectively provide support for the security protection of private data sets.

Keywords：Rivest Shamir Adleman(RSA) algorithm；Data Encryption Standard(DES) algorithm；hybrid algorithm； privacy data set； homomorphic encryption

在计算机技术与网络技术迅速发展的背景下，信息数字化程度也在不断提高，大数据量呈现出急剧增长的态势[1]。这其中就包含了大量的用户隐私数据。隐私数据是指具有私密性、安全防护等级高的敏感性强的数据。数据在存储和检索处理过程中最容易被攻击和窃取[2-3]。因此，数据隐私的安全保护成为维护数据安全、平稳运行中亟待解决的问题，更是信息安全领域中的重要课题。张文娟等[4]将多值与模糊方案相结合应用至数据安全保护中，证明算法可以高效保障云端和云辅助无线体域网之间数据通信安全性，并与当前加密算法相结合，给出一种实现算法的实际示例，表明该算法可在一定程度上保障数据安全性。赵志远等[5]针对物联网中的数据安全问题，提出属性基加密法。该方法在保障数据隐私前提下，可以实现密文数据细粒度访问管控。

收稿日期：2020-03-01；**修回日期：**2020-05-08

基金项目：山西省高校教学改革创新资助项目(J2018178)；晋中学院“1331 工程”创客团队建设计划基金资助项目(jzxycktd2017018)

作者简介：鲍海燕(1982–)，女，硕士，讲师，主要研究领域为网络安全。email:522219898@qq.com

此外，加密方法还包括应用非对称密钥(RSA)算法、数据加密(DES)算法和Noekeon 算法等。然而，在应用传统算法或只应用单项算法实现数据加密的过程中，难以有效抵御多变的攻击或窃取行为，导致加密效果理想度较差。同态加密不仅具有非常高的安全性，且能够在加密之后的数据上直接进行处理，该方法下的解密结果和直接对明文进行处理所得结果相同，可实践性非常强。为此，结合当前数据加密相关研究成果，本文对传统 RSA 算法进行改进优化，并将其与传统 DES 算法结合组成混合算法，共同实现对隐私数据集的同态加密处理。

1 隐私数据集同态加密方法设计

1.1 RSA 算法及其改进

1) RSA 算法加、解密过程：

步骤 1：选取 2 个大素数 a 和 b，且 $a \neq b$， $f = a \times b$， $\phi(f) = (a-1) \times (b-1)$；

步骤 2：选取素数 g，使 g 满足 $1 < g < f$；

步骤 3：将 (g, f) 作为公钥，对明文实行加密操作；

步骤 4：对私钥 $h = \dfrac{\mathrm{mod}\,\phi(f)}{g}$ 进行计算；针对密文实行解密操作。

上述步骤中，g 和 f 都是公开的，但 $\phi(f)$ 是密钥，需要保密。$\phi(f)$ 被获取后，该 RSA 算法安全性会大大削弱。RSA 算法的安全性与 a 和 b 长度存在直接关联性，a 和 b 的长度越长，安全性就越高。因此，实际应用 RSA 算法的过程中，a 和 b 的长度至少为 512 bit。

2) 参数选择

RSA 算法是第 1 个把系统安全性依托在因数分解上的加密算法[6]。如果 f 非常容易被分解，则 RSA 算法的安全性将大大降低。综上，能够认为 RSA 算法的安全性与因数分解为等价关系。即：在 RSA 算法运行过程中，对于公钥的选择非常重要。

参数 f 的选择原则：在该过程中，f 值要足够大，此为保障 RSA 算法安全性能最根本的原则。为保障算法安全性，产生的大素数长度均至少为 100 位之上的十进制数，这时的 f 能够在 200 位以上[7]。

在上述环境下，计算 $f = a \times b$，a 和 b 最好均为强素数。其中，强素数可定义为：

条件一：有大素数 a_1 和 a_2，可以使 $a_1 \mid (a-1)$ 和 $a_2 \mid (a+1)$；

条件二：有强素数 r_1，r_2，s_1 和 s_2，可以使 $r_1 \mid (a_1 - 1)$，$s_1 \mid (a_1 + 1)$，$r_2 \mid (a_2 - 1)$ 和 $s_2 \mid (a_2 + 1)$。

条件三：a 和 b 之间的差要很大，同时 $a-1$ 和 $b-1$ 的最大公因子要足够小。假设 a 和 b 之间差值很小，则可使 $(a+b)/2 \approx \sqrt{f}$ 成立，能够计算得到 a 和 b 值。

参数 g 的选择原则：在 RSA 算法运行过程中，g 仅需满足 $\gcd(g, \phi(f)) = 1$，则表示 g 能够随机选择。根据加密的基本原理可知，g 值越小，加密需要的时间就越少。因此，g 越小，算法运行效果越好。但实践表明，g 值小，会触发安全方面的问题[8]。综上，实际运行中，参数 g 选择需要遵循的条件如下：

条件一：g 不可过小，综合考虑数据加密的效率和安全性，最好选取长度为 16 的素数。

条件二：参数 g 选取过程中，应该选取模 $\phi(f)$ 阶最大，也就是 $g^i = 1 \bmod \phi(f)$ 中最小的 i 应为 $((a-1)(b-1))/2$。

参数 h 的选择原则：h 作为密钥应该大于 $f^{1/4}$，实际应用过程中，均希望根据小位数 h 提高解密或签名效率。当 g 确定之后，可基于 Euclidean 算法获取 h。综上，假设 h 长度小于 $f^{1/4}$，则仅需通过数学算法就能够高效率。求得参数 h。

3) 改进 RSA 算法

RSA 算法具备的优点和缺点均较为显著，其优点为安全性能高，但加密的效率较低，耗时较久[9]。在 RSA 算法运行过程中，素数生成和判定、模幂计算较为耗时。因此，在 RSA 算法改进阶段，主要针对大素数生成和判定、模幂计算过程进行优化。

素数判断过程已在上文中进行了分析，下面主要针对模幂的快速计算进行讨论。

模幂的快速计算需采用滑动窗口法，滑动窗口法的原理为：把模幂 $c^{g'} \bmod f$ 中的指数 g' 实行模块化操作。假设 $e' = (15454855115)_{10} = (0011100110010010111001011111111001011)_2$，设定窗口的长度为 3，将上述的 2 进制数由左边起分组，且分组的长度为 3 位。分组原则为保障每组第一位是 1，由此中间可能会将部分 0 跳过[10]。在最后当长度比 3 位数小，则将剩下的数单独构成一组，无需补 0 或补 1[11]。

分组后，对分组提高模幂运算效率的方式进行分析。以 $p^{(1011)_2}$ 为例：$p^{(1011)_2}=p^{11}=p^{10}p=p^{(101)_2\times 2}p$。由此能够推导出 g' 的表现形式：

$$g'=7\times 2^{31}+6\times 2^{26}+4\times 2^{22}+5\times 2^{19}+6\times 2^{16}+5\times 2^{12}+7\times 2^{9}+7\times 2^{6}+5\times 2^{1}+2^{0} \tag{1}$$

简化 g'，同时将简化之后的 g' 代入模幂 $c^{g'}\bmod f$ 中。

根据上述计算和分析可知，滑动窗口法运行的本质即为对指数实行预处理，以此减少计算量。

1.2 隐私数据集同态加密的实现

1) 算法混合

RSA 算法虽然有效性得以提高，但加密速度有待提升，为此引入 DES 算法。DES 算法加密速度较快，但密钥管理存在一定问题[12]。因此结合改进 RSA 算法和 DES 算法，能够实现有效互补。步骤如下：

步骤 1：定义：

$$F_{\mathrm{DES}}=P'^{-1}\times T_{16}\times T_{15}\times\cdots\times T_{1}\times P'(MING) \tag{2}$$

式中：P' 为初始变换，P'^{-1} 为 P' 的逆变换，两者满足 $IP'\times IP'^{-1}=1$；T 为编码操作轮数。

步骤 2：DES 迭代；

步骤 3：在子密钥生成过程中，将原来密钥进行重新排序，划分为两部分，然后通过循环移位获取子密钥的两部分，最后合成并重新排序形成子密钥。混合加密后的解密过程与加密步骤类似，在解密过程中把 16 轮迭代自密钥顺序倒过来即可。

在混合算法运行过程中，通过 DES 算法对明文数据集进行加密，并对 DES 算法用到的密钥实行 RSA 加密。将打包密文与加密后的密钥传输至接收端，接收端获取数据包后，对密钥密文实行解密操作，得到 DES 算法加密时的密钥。DES 算法加密与解密所用密钥相同，在得到密钥后即能实现密文解密。图 1 为混合算法运行原理。

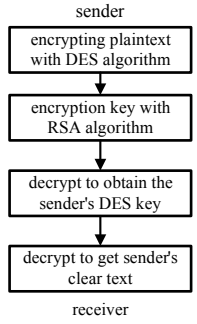


Fig.1 Flow of hybrid algorithm

图 1 混合算法运行原理

在 VC++平台上实现混合算法。加密文件为.txt 格式，首先根据 DES 算法实现文件加密,并针对 DES 算法密钥实行 RSA 加密,利用微软基础类库(Microsoft Foundation Classes，MFC)界面显示加密结果。此过程中，RSA 算法 a 和 b 值均为 512 bit。将加密过程中消耗的时间，包括 RSA 算法密钥生成时间全部在界面上进行实时显示。文件解密过程中，先要得到 DES 算法密钥。由此，解密操作中，首先对 RSA 加密的 DES 密钥进行解密，并输出结果，然后根据该密钥对文件实行 DES 解密。

2) 同态运算

在实现对隐私数据集混合加密的基础上，通过同态运算，实现隐私数据集加密。

同态运算是针对密文进行设计的一种加密过程，根据对密文的运算结果，可以得到相应的明文结果。本文设计的同态运算过程包括：密钥生成计算、加密、解密和密文同态计算。其中，密钥生成计算、加密和解密过程已在上文描述，本节主要对密文同态计算过程进行分析。

选择加法同态实现对密文的同态计算。过程如下：a) 利用输入素数生成密钥；b) 利用密钥加密明文，返回密文；c) 利用密钥解密密文，返回明文；d) 假设 C 代表明文空间中的一个运算集合，记为 $C\{+\}$。对于输入的数据 $c_n(n=1,2,3,\cdots)$，首先对数据 c_n 进行加密，并将其转换至密文空间中，得到密文运算结果。用 $\oplus$ 表示加法同态处理，则 $((c_1,c_2),+)=c_1\oplus c_2$，$((c_2,c_3),+)=c_2\oplus c_3$，以此类推。通过对密文的计算以及对结果的解密操作，可以得到相应的明文计算结果。

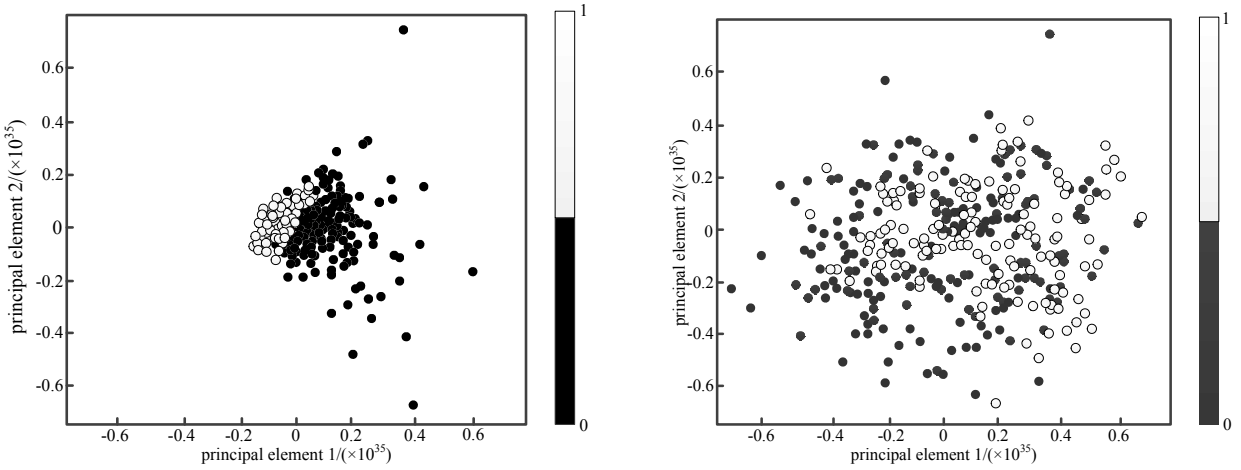
综上所述，实现了基于改进 RSA 算法的隐私数据集加法同态加密。

2　实验结果与分析

为验证基于改进 RSA 算法的隐私数据集同态加密方法的有效性，进行如下性能测试。

实验搭建于 Linux 平台中，主机配置双核 4.0 GHz 主频的 CPU 和 128 GB 内存。开发语言为 VC++，实验数据为大小 10 GB 的.txt 文本文件，实验数据均来自 Github 数据集。

首先，对数据集的加密性能进行初步验证。对比加密前后的数据维度，结果见图 2。分析图 2 可知，在加密前数据维度较为集中，但应用基于改进 RSA 算法的隐私数据集同态加密方法加密后，所显示出来的数据维度较为分散，有效保证数据信息的私密性，初步证明了本文方法的有效性。



(a) before encryption (b) after encryption

Fig.2 Data dimension comparison before and after applying encryption method

图 2 应用加密方法前后数据维度对比

为进一步验证本文方法的有效性，对本文方法、传统 RSA 算法加密方法和 DES 算法加密方法的加密过程耗时进行对比，结果见图 3。分析图 3 可知，传统 RSA 算法的加密方法的加密过程耗时接近 10 s，基于改进 RSA 算法的加密方法的加密过程耗时在 5~6 s 之间，DES 算法的加密方法的加密过程耗时在 3~4 s 之间。通过对比可知，本文所设计的同态加密方法的加密过程耗时仅略多于 DES 算法，说明基于改进 RSA 算法的加密方法具有较高的加密效率。

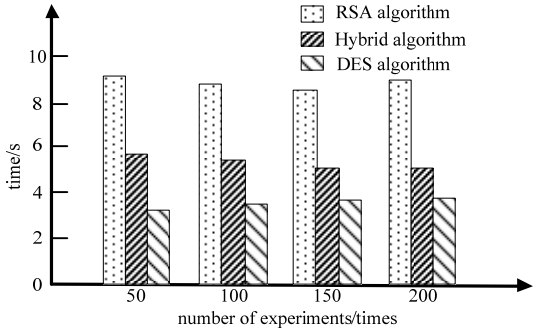


Fig.3 Time consumption of encryption process with different encryption methods

图 3 不同加密方法加密过程耗时对比

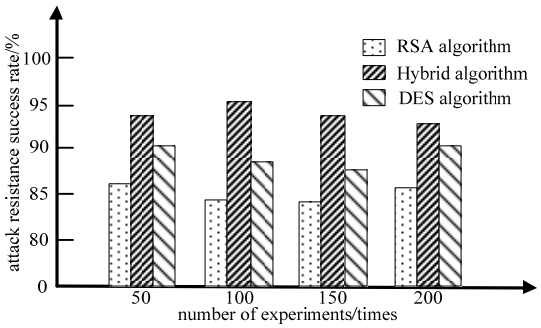


Fig.4 Success rate of different encryption methods against attacks

图 4 不同加密方法抵御攻击成功率对比

在此基础上，在加密传输过程中，对隐私数据集实施篡改攻击，并测试不同方法的抵御攻击成功次数，从而判断不同方法的抵御攻击成功率，对比结果见图 4。分析图 4 可知，传统 RSA 算法的加密方法的抵御攻击成功率在 85%左右，DES 算法的加密方法的抵御攻击成功率在 90%左右，而基于改进 RSA 算法的加密方法的抵御攻击成功率在 3 种方法中为最高，保持在 95%左右。由此可知，与单项算法相比，混合算法抵御攻击成功率更高，更能保障数据集的安全性。

表 1 为不同数据集加密法整体性能，通过该表可更为明显地反映不同数据集加密法性能的优越性。根据表 1 可知，相比其他加密算法，本文方法应用混合算法后的综合性能更为优越，很好地弥补了 DES 算法和 RSA 算法的缺陷，不仅提高了数据集加密效率，还增强了数据集的安全性。

表 1 不同数据集加密法性能对比

Table1 Performance comparison of encryption methods for different data sets

performance indicators	comparing the results
encryption efficiency	DES algorithm ≈ mixing algorithm >RSA algorithm
safety performance	hybrid algorithm >RSA algorithm >DES algorithm
key generation and control	hybrid algorithm >DES algorithm ≈RSA algorithm
implementation and application	compared with single-item algorithm, hybrid algorithm has a wider application range and environment

3　结论

为保护用户隐私数据的安全性，本文在改进RSA算法的基础上，将其与DES算法结合起来，并设计加法同态计算实现对隐私数据集的加密。实验结果证明了该方法具有较强的加密性能，可实践性更强。研究中发现：在改进RSA算法中，RSA算法的安全性与因数分解有关，在选择公钥素数时，素数值越小，加密过程需要的时间就越少。因此，素数越少，则越能提高算法的运行效率。随着计算机技术的日益发展，隐私数据入侵技术也在不断精进，数据加密方法将会受到更大的挑战。在接下来的研究中，要不断对加密方法进行改进，与相关硬件系统相结合，为保障数据安全性奠定更为坚实的基础。

参考文献：

[1] 王煜,朱明,夏演. 非对称加密算法在身份认证中的应用研究[J]. 计算机技术与发展, 2020,30(1):94-98. (WANG Yu, ZHU Ming,XIA Yan. Application research of asymmetric encryption algorithm in identity authentication[J]. Computer Technology and Development, 2020,30(1):94-98.)

[2] 刘海峰,刘洋,梁星亮. 一种结合优化后 AES 与 RSA 算法的二维码加密算法[J]. 陕西科技大学学报, 2019, 37(6):153-159. (LIU Haifeng,LIU Yang,LIANG Xingliang. A QR code encryption technique combining optimized AES and RSA algorithms[J]. Journal of Shaanxi University of Science & Technology, 2019,37(6):153-159.)

[3] 柯彦,张敏情,刘佳. 可分离的加密域十六进制可逆信息隐藏[J]. 计算机应用, 2016,36(11):3082-3087. (KE Yan, ZHANG Minqing,LIU Jia. Separable reversible hexadecimal data hiding in encrypted domain[J]. Journal of Computer Applications, 2016,36(11):3082-3087.)

[4] 张文娟,吴聪,余梅生,等. 利用多值和模糊属性的云辅助 WBAN 数据加密算法[J]. 计算机应用研究, 2016,33(5): 1537-1541. (ZHANG Wenjuan,WU Cong,YU Meisheng,et al. Data encryption algorithm of cloud-assisted WBAN using multi-valued and ambiguous attribute[J]. Application Research of Computers, 2016,33(5):1537-1541.)

[5] 赵志远,王建华,朱智强,等. 面向物联网数据安全共享的属性基加密方案[J]. 计算机研究与发展, 2019,56(6): 1290-1301. (ZHAO Zhiyuan,WANG Jianhua,ZHU Zhiqiang,et al. Attribute-based encryption for data security sharing of internet of things[J]. Journal of Computer Research and Development, 2019,56(6):1290-1301.)

[6] 庞金香,隋萌萌. 同态加密隐私保护数据高效智能挖掘仿真[J]. 计算机仿真, 2019,36(6):316-319. (PANG Jinxiang, SUI Mengmeng. Homomorphic encryption privacy protection data efficient intelligent mining simulation[J]. Computer Simulation, 2019,36(6):316-319.)

[7] 黄峰,邓辉明,白世雄. 基于 CPS 的网络控制系统双重信息加密算法设计[J]. 湖南工程学院学报(自然科学版), 2019,29(1):1-4. (HUANG Feng,DENG Huiming,BAI Shixiong. A design of dual information encryption algorithm for network control system based on CPS[J]. Journal of Hunan Institute of Engineering(Natural Science Edition), 2019,29(1):1-4.)

[8] 杨正文,郭筝. 抗相关性功耗分析的 DES 掩码方案[J]. 微电子学与计算机, 2019,36(9):1-6. (YANG Zhengwen,GUO Zheng. Masking scheme against correlation power analysis on DES[J]. Microelectronics & Computer, 2019,36(9):1-6.)

[9] 周金治,高磊. 基于多素数和参数替换的改进 RSA 算法研究[J]. 计算机应用研究, 2019,36(2):495-498. (ZHOU Jinzhi,GAO Lei. Research on improved RSA algorithm based on multi-prime number and parameter substitution[J]. Application Research of Computers, 2019,36(2):495-498.)

[10] 王爽,陈丽. DES 解密过程与安全性的探讨[J]. 新疆师范大学学报(自然科学版), 2019,38(1):39-42. (WANG Shuang, CHEN Li. Discussion of DES decryption process and security[J]. Journal of Xinjiang Normal University(Natural Sciences Edition), 2019,38(1):39-42.)

[11] 徐鹏,薛伟. 抗差分功耗攻击的 DES 算法研究[J]. 计算机仿真, 2018,35(1):282-286. (XU Peng,XUE Wei. Research of DES algorithm against differential power analysis[J]. Computer Simulation, 2018,35(1):282-286.)

[12] 薛燕,朱学芳. 基于改进加密算法的云计算用户隐私保护研究[J]. 情报科学, 2016,34(9):145-149. (XUE Yan,ZHU Xuefang. Research on improved encryption algorithm for privacy-preserving problems under cloud computing environment[J]. Information Science, 2016,34(9):145-149.)