一种基于前向纠错码的图像 DNA 加密存储算法

姚翔宇[1], 苏燕青[1], 昝乡镇[1], 许 鹏[1], 刘文斌[1]

[1]广州大学计算科技研究院 广州 中国 510006

摘要 近年来 DNA (DeoxyriboNucleic Acid) 存储发展迅速, 实现数字图像 DNA 存储和安全传输成为有待解决的问题。因此该文提出了一种面向 DNA 存储的基于前向纠错码的图像加密算法。首先使用动态约瑟夫遍历算法对图像像素点进行行置换和列置换, 以消除明文图像相邻像素之间的相关性。其次, 使用图像分解方法将明文图像分解为 8 个子图, 然后再重新组合, 实现了对图像像素值的置换, 从而进一步消除明文图像的纹理特征和破坏其统计学特征。再次, 对图像进行全局扩散, 使明文的微小变化以扩散的形式影响密文, 以抵抗差分攻击。最后使用可纠错 DNA 编码表将图像加密编码为 DNA 序列, 合成后进行存储。算法将明文图像加密成 DNA 序列并存储, 这种存储方式与传统存储介质相比更为安全。同时, 可纠错 DNA 码使得密文可以在 DNA 存储环境中可靠读取。该文使用 3 张常用图像包括 lena_gray、peppers_gray、baboon_gray, 测试算法的安全性以及在 DNA 存储环境下的鲁棒性。仿真结果表明, 该方法可以有效抵御多种密码学攻击, 并且在 DNA 存储环境下对碱基错误和序列缺失等问题表现出良好的鲁棒性。

关键词 图像加密; 约瑟夫遍历; 图像分解; 可纠错 DNA 码

中图法分类号 TP309.7 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2023.11.03

An Image Encryption and Storage Algorithm Based on Forward Error Correction DNA Codes

YAO Xiangyu[1], SU Yanqing[1], ZAN Xiangzheng[1], XU Peng[1], LIU Wenbin[1]

[1] Institution of Computational Science and Technology, Guangzhou University, Guangzhou 510006, China

Abstract In recent years, DNA (DeoxyriboNucleic Acid) storage has developed rapidly, the storage of digital image in DNA and its security of transmission has become a problem to be solved. In this paper, an DNA storage-oriented image encryption storage algorithm based on forward error correction code is proposed. Firstly, we apply dynamic Joseph traversal to scramble the rows and columns of the plaintext image. Secondly, the image decomposition method is used to decomposite the plaintext image into 8 subgraphs, thus further eliminating the texture features of the plaintext image and transforming the statistical distribution of pixel values. Thirdly, the image is globally diffused, so that the little changes in plaintext image will affect the whole ciphertext in the form of diffusion, which make the alogrithm has the ability to resist differential attack. Finally, the error correcting DNA coding table is used to encrypt and encode the image into DNA sequences, which are synthesized and stored after encrytion. In this way, the plaintext image is encrypted in DNA sequences which consist of mang error corretion DNA codes. Compared to traditional image encryption algorithm, the ciphertext of the proposed encryption algorithm is storaged in DNA storage system rather than traditional storage device which is easier to be cracked. Meanwhile, by using the error correction DNA codes in encryption process, the ciphertext of the proposed algorithm can reliablely reading and writing in DNA storage system. In this paper, a general image set including lena_gray, peppers_gray and baboon_gray is used to test the security and robustness in DNA storage system of the proposed image encryption algorithm. Simulation results show that the proposed image encryption algorithm can effectively resist various attacks and present a high robustness in against base error and sequence loss which are the specific problems of the current DNA storage system.

Key words image encryption; joseph traversal; image dissection; DNA error correction codes

1 引言

数字图像是当今广泛使用的多媒体数据, 特别在医疗, 商业, 军事等方面起着重要的作用。为实现数字图像的安全传输, 高安全性的图像加密技术成为多媒体通信领域中最令人关注的问题之一。近年来, 由于混沌系统具有简单易实现、对初值敏感、不可预测、非周期等特性, 研究者们提出许多基于混沌

通讯作者: 刘文斌, 博士, 教授, Email: wbliu6910@gzhu.edu.cn。

本课题得到国家自然科学基金(No. 62072128, No. 61876047, No. 62002079)资助。

收稿日期: 2022-03-22; 修改日期: 2022-05-25; 定稿日期: 2023-09-01

系统的图像加密算法。其中一些算法实现了一次一密, 还有一些算法具有较大的密钥空间足以抵抗暴力破解。随着存储技术和密码学的发展, 科学家们发现 Deoxyribonucleic Acid(DNA)是一种理想的数字文件存储介质[1-6], 因 DNA 有高并行性及高存储密度等优点, 基于 DNA 的加密方法成为了传统加密技术的潜在替代途径[7-9]。目前, DNA 存储及 DNA 密码正引起越来越多的研究者关注。

早在 1999 年, Celland 将二战中著名的军事密函“June 6 Invasion: Normandy”隐藏在 DNA 微点中[10], 这是科学界首次实现用 DNA 加密信息。2003 年, Gehani 等人[11]借助 DNA 作为信息载体, 利用生化技术在 DNA 分子上实现了一次一密的传统加密算法。2007 年至 2010 年, 卢明欣等人[12-13]提出了基于 DNA 技术的对称与非对称加密方法(DNASC/DNAPKC), 能够抵御超级计算机的攻击。

目前主流利用 DNA 进行加密的方法可分为两种: 1) 使用 DNA 数据库作为密码本构建一次一密加密系统。2015 年, 王兴元等人[14-15]利用碱基加、减、互补等运算规则进行矩阵运算, 最后将 DNA 矩阵转化成密文图像完成加密。2017 年, Thangavel 等人[16]利用二进制文件在 DNA、mRNA、tRNA、特殊字符之间的相互转换完成加密和解密。2020 年, 张勋才等人[17]利用 DNA 动态编码以及 GenBank 数据库实现图像加密。2) 利用生物技术和处理 DNA 的困难问题构建全新的 DNA 密码系统。2016 年 Zakeri 等人[18]利用 DNA 序列中短串联重复序列的丰富数量生成高强度密钥进行加密。2019 年, 上海交通大学樊春海等人[19]使用 DNA 折纸术[20]加密信息。2021 年, Shuang Cui 等人[21]在 DNA 密文序列中添加特定的引物控制 DNA 测序, 进而实现保密功能。

由于 DNA 存储系统会出现存储序列缺失以及碱基插入, 删除, 替换错误, 基于 DNA 存储的图像加密算法解密难度大且成功率低。因此我们提出一种面向 DNA 存储的基于前向可纠错码的图像加密算法。算法包括基于混沌系统的像素置乱和扩散, 基于图像分解的像素置换, 以及可纠错 DNA 编码表。仿真实验结果表明, 该方法安全性较高, 并且能够在 DNA 存储环境下成功解密。

2 相关理论

2.1 Rossler 混沌系统

Rossler[22]模型是一种经典的混沌系统, 其动力学方程组形式如(1)式。

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=-\omega y-z\\ \dfrac{\mathrm{d}y}{\mathrm{d}t}=\omega x+\alpha y\\ \dfrac{\mathrm{d}z}{\mathrm{d}t}=\beta+z(x-\gamma)\end{cases}\tag{1}$$

选取合适的参数值(α, β, γ, ω),该系统可产生多条混沌序列。

2.2 Keccak 算法

2012 年 10 月, 美国国家标准与技术研究院选定 Keccak[23]作为新哈希函数标准算法。对于输入的明文图像, Keccak 算法输出“图像指纹”。哈希算法是不可逆的, 因此无法通过“图像指纹”还原出明文图像。

2.3 约瑟夫遍历

约瑟夫遍历可用函数$f(S, l)$表示, S表示序列长度, l 表示遍历步长。例如 $f(9, l)$, l=(2, 4, 1, 7, 5, 6, 9, 3, 8)。它表示将序(1, 2, 3, 4, 5, 6, 7, 8, 9)首尾相接排成环状, 从 1 开始, 步长 2, 遍历的第一个数是 2, 然后从 3 开始, 步长 4, 遍历第二个数为 6。如此循环直至遍历完所有数字。最终遍历顺序为 2, 6, 7, 8, 5, 1, 9, 3, 4。

2.4 可纠错 DNA 编码表

为纠正 DNA 存储系统的碱基错误, 我们在算法中设计了一种可纠错 DNA 编码表(表 1)。它包含 64 个 6 碱基编码, 可以编码 6 比特信息。为满足 DNA

表 1 编码表

Table 1 Coding table

DNA 编码			
TAACCG	CTACCA	CTTAAC	GCCATA
GATTGA	TACGCT	CAATTG	TCGGCG
ATAGTG	AGAAGA	TATGAG	CTGTGC
TGACAC	TGAGGT	CACAAT	AGTAAT
TCAACT	GTATTC	ACTGGT	GCAGGA
AGCATC	CGCTTA	CTTGTA	GAATAT
CTCTCG	CTGACT	GACCTG	TCCGTC
TAGCGA	GTCCGA	TCATGC	CGGCTC
CGTCGT	TGCTAT	ACGCCA	TGGCCT
TCTAGA	TAAGTA	GTTGCT	ACATTA
ATTCGC	CAGATA	GCGTTG	ATTACA
AGGTCG	ACCTCT	GCGAGT	GACGGC
CAGCAG	ACGAAC	GTGGAC	CGGGCA
GTGTCA	CAAGAC	GAGACG	CAGGGT
CATTCT	TGGATG	ACTCAG	CGAAAG
GTCAAG	ATCCAT	AGTCTA	ATGGGA

存储系统及序列纠错的要求, 编码表具有如下特点: 1) 编码的总体 GC 含量基本分布均匀; 2) 任意编码拼接不会产生长度大于 3 的均聚物; 3) 任意两个编码之间的编辑距离至少为 3。

3 加密与解密

3.1 密钥生成

我们把明文图像输入Keccak函数, 输出记为 K, 输出长度为 512。然后将 K 均匀分成 32 组, 即 $K=\{k_1, k_2, k_3, \cdots, k_{32}\}$。最后按照(2)式和(3)式计算密钥 x_0, y_0, z_0,计算精度为小数点后 16 位。

$$h_i=(k_i+k_{i+5}+k_{i+10}+k_{i+15}+k_{i+20}+k_{i+25})/256 \quad (2)$$

$$\begin{cases} x_0=h_1 \bmod 1 \\ y_0=h_2 \bmod 1 \\ z_0=h_3 \bmod 1 \end{cases} \quad (3)$$

我们将 x_0, y_0, z_0 作为初值输入 Rossler 系统, 系统随即产生三条混沌序列 x, y, z。我们用(4)式得到 x', y', z'。式中 M, N 分别表示明文图像的行和列。

$$\begin{cases} x'=(x*10^{16}) \bmod M+1 \\ y'=(y*10^{16}) \bmod N+1 \\ z'=(z*10^{16}) \bmod 64 \end{cases} \quad (4)$$

3.2 置乱

置乱步骤如下。

步骤一: 将序列 x', y' 以列优先方式排成与明文图像大小相等的矩阵 $\boldsymbol{X}$, $\boldsymbol{Y}$。

步骤二: 矩阵 $\boldsymbol{X}$ 的第 i 行作为图像第 i 行的约瑟夫遍历步长, 对图像进行行置乱。

步骤三: 矩阵 $\boldsymbol{Y}$ 的第 i 列作为图像第 i 列的约瑟夫遍历步长, 对图像进行列置乱。

下文用 P 表示将明文图像, P_1 表示置乱后的图像。

3.3 基于图像分解的像素置换

具体步骤如下。

步骤一: P_1 每个像素包含 8 比特信息, 取出第 i 个比特组成 8 个大小为 $M\times N$ 的 0,1 矩阵 m_i。

步骤二: m_i 以行优先排列成大小为 $1\times MN$ 的矩阵。

步骤三: 将矩阵从左至右每 6 个元素转化成一个十进制数字, 得到长度为 $MN/6$ 的序列。

我们令 $S=(s_1, \cdots, s_i, \cdots)$。$s_i$ 为步骤三得到的序列, $i\in[1,8]$。如果明文图像所含的比特数总量不能被 6 整除, 我们用值为 0 的像素点对图像进行行或者列的填充。

3.4 全局扩散

函数定义: sort(a,b)=b_0 表示将序列 a 以升序排序, 序列 b 以 a 为基准重新排序得到 b_0。如 a=(3,4,2), b=(e,c,d), sort(a,b)=(d,e,c)。

首先从混沌序列 z 的 r 位置向后截取一段长度为 8 的序列, 记作 e。然后 S= sort(e, S), 并把 S 中的序列串联成一条长为 $4MN/3$ 的序列。接下来按(5)式进行扩散。式中 S_i 表示 S 的第 i 个元素, z_i' 表示 z' 的第 i 个元素。

$$\begin{cases} S_i'=S_1\oplus S_{4MN/3}\oplus S_{4MN/3-1}\oplus z_1', (i=1) \\ S_i'=S_2\oplus S_1'\oplus S_1\oplus z_1', \quad (i=2) \\ S_i'=S_{i-1}'\oplus S_{i-2}'\oplus S_i\oplus z_i', \quad (i>3) \end{cases} \quad (5)$$

最后我们把 S' 的每个元素转化成 6 比特的二进制数并串联在一起, 然后从左至右每 8 比特转化成一个十进制数, 以行优先排成大小为 $M\times N$ 矩阵, 得到密文图像 C。

3.5 DNA 编码加密图像

我们使用表 1 加密 C。具体步骤如下。

步骤一: 以列优先将表 1 中 DNA 码排列成序列 D。

步骤二: 从混沌序列 z 的 t 位置向后截取一段长度为 64 的序列, 记作 g。然后 D= sort(g, D)。

步骤三: 把 C 每个像素点转化成 8 位二进制数并串联在一起。D 中编码按先后顺序与 000000-111111 形成一对一映射, 将 C 转化为 DNA 序列。

该算法同样适用于彩色图像, 将彩色图像 RGB 三通道分离处理即可。加密流程如图 1 所示。由于算法是对称及可逆的, 所以解密为加密的逆过程。密钥为 x_0, y_0, z_0, r, t。解密前需用文献[24]或文献[25]的 DNA 序列纠错算法对密文序列纠错。

4 仿真实验及结果分析

4.1 密钥空间分析

由于 Rossler 系统三个初值的计算精度, 像素置换中 8 个分解图像的排序方式, 以及 DNA 编码的排序方式, 密钥空间以如下方式计算:

$$10^{16}\times 10^{16}\times 10^{16}\times 8!\times 64!\geq 2^{128*3} \quad (6)$$

其中, 计算机的精度为 10^{-16}, 分解图像的排序方式共 8! 种, DNA 码排序方式共 64! 种。该算法的密钥空间远大于 2^{128}, 因此可抵抗暴力破解。

4.2 直方图分析

直方图是通过统计像素的频数而得到的函数图,

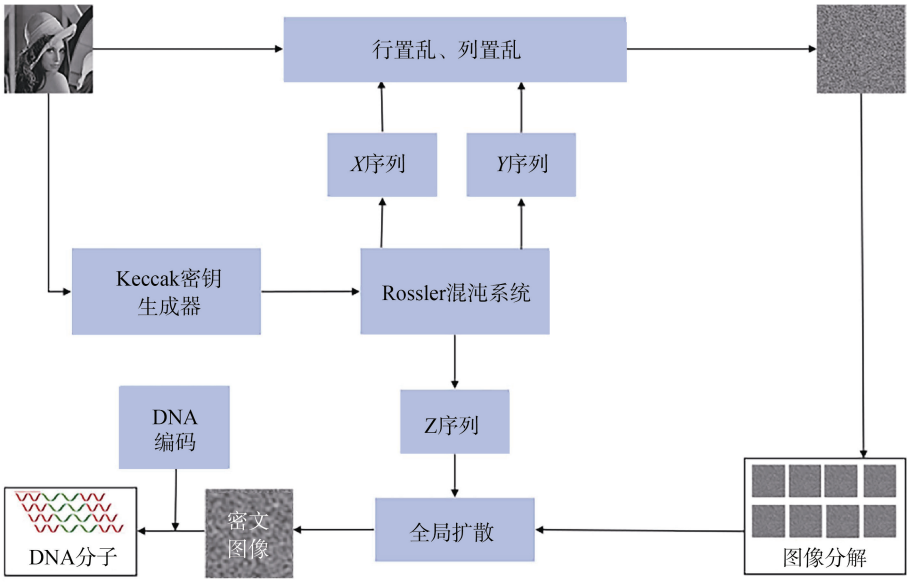


图 1　加密流程图

Figure 1　Encryption flow chart

可评估算法抗统计分析能力。本方案将图像加密为 DNA 序列, 因此我们统计明文图像像素点直方图和密文序列中 DNA 码直方图。

图 2 是 lena_gray 和 peper_gray 的明文和密文直方图。从图中可以看到, 密文直方图较明文直方图平整, 并且密文序列中 DNA 码频数分布十分均匀。因此该方案具有很好的抵抗统计学分析能力。

直方图的分布规律可用 χ^2 分布来衡量。因此本文给 lena_gray 和 peppers_gray 的明文和密文的直方图做了 χ^2 测试, 显著性水平 α =0.05。我们用(7)式计算 χ^2。式中 w_i 表示直方图的列。

$$\chi^2 = \frac{1}{64}\sum_{i=0}^{63}(w_i - \frac{1}{64}\sum_{i=0}^{1}w_i)^2 \tag{7}$$

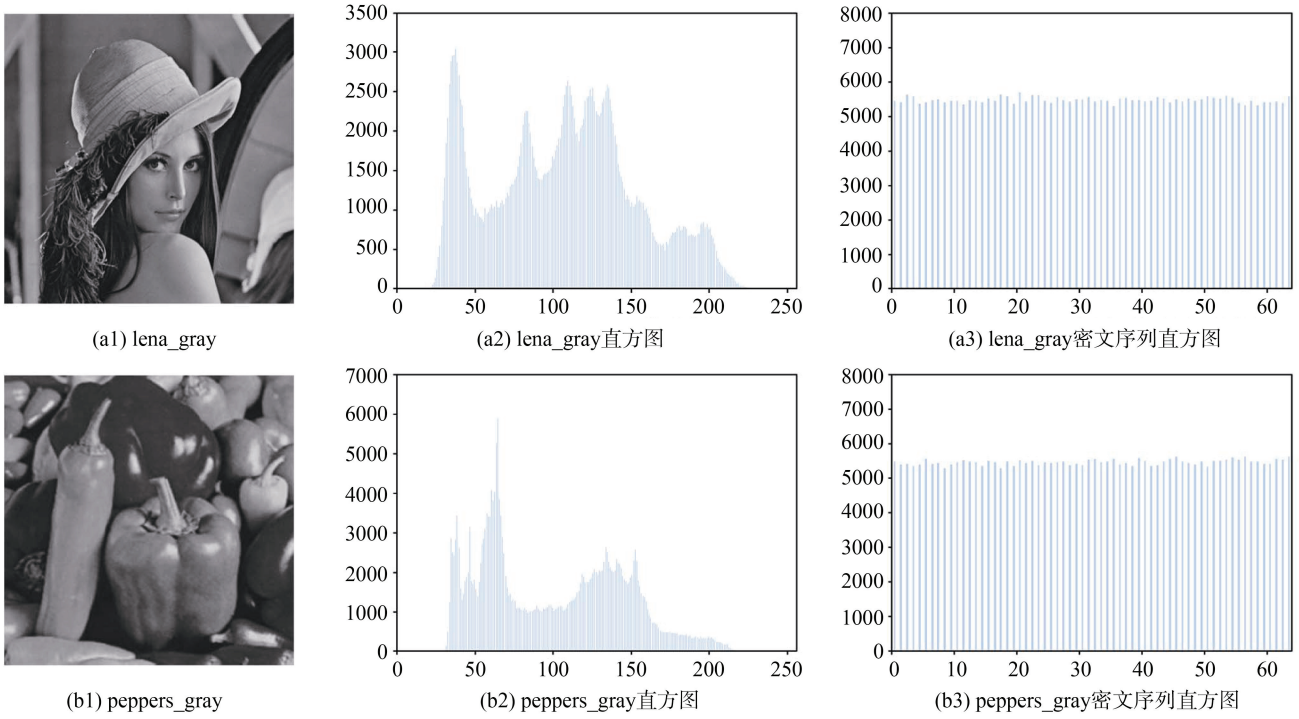


图 2　直方图

Figure 2　Histograms

表 2 为 χ^2 检测结果。如表中所示, lena_gray 和 peppers_gray 的密文直方图 χ^2 分别为 53.1835 和 60.8242, 小于 $\chi^2_{0.05}(63)$。因此密文直方图服从自由度为 63 的 χ^2 分布。而明文直方图 χ^2 大于 $\chi^2_{0.05}(63)$。由此可知该方法能够有效破坏原始图像的统计特征。

表 2 直方图 χ^2 测试结果

Table 2 χ^2 text results

	明文直方图 χ^2	密文直方图 χ^2	$\chi^2_{0.05}(63)$	检测结果
lena_gray	157.6657	53.1835	82.5287	通过
peppers_gray	204.3332	60.8242	82.5287	通过

4.3 相关性分析

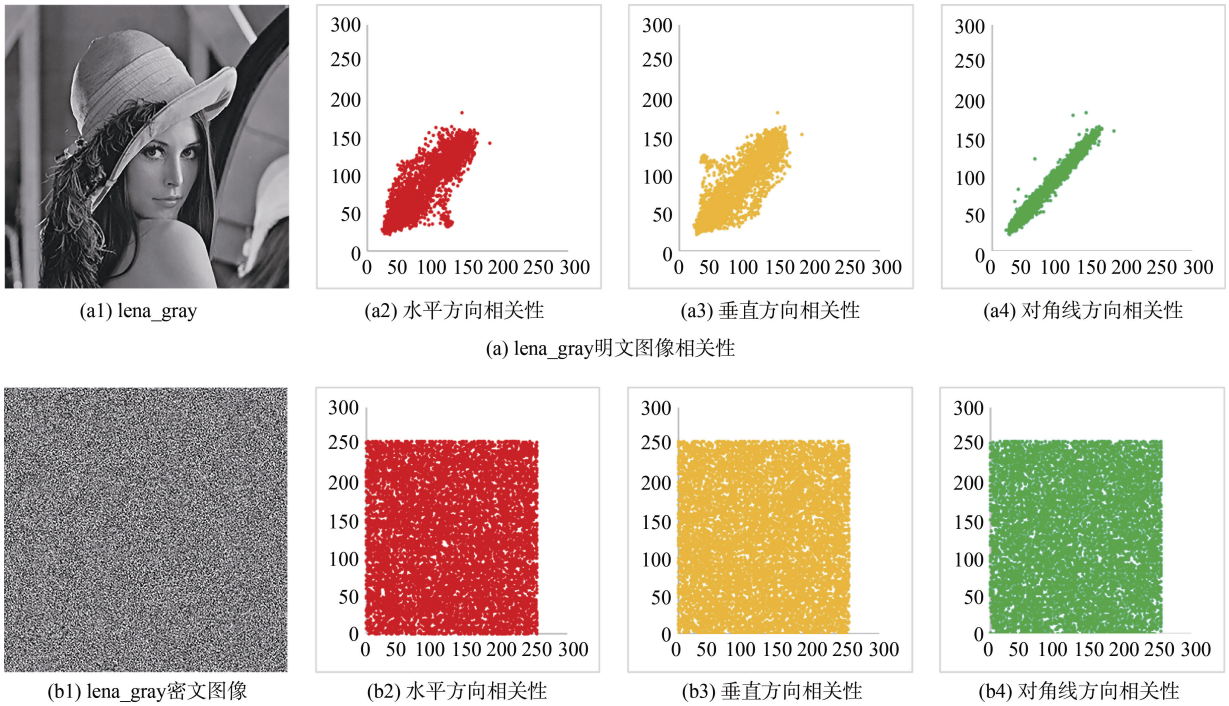
除了直方图分析之外, 还有一种统计学分析为图像相邻像素点的相关性分析。我们在 lena_gray 的明文图像和密文图像中随机选取 10000 个像素点, 从水平、垂直、对角线三个方向进行相关性分析。

相关性分析结果如图 3 所示。从图中可以看到, 明文图像相邻像素值分布大致呈线性, 说明像素间具有很强的相关性。而密文图像相邻像素值分布呈现随机性, 说明像素之间相关性较小。

本文用(8)式和(9)式计算相关性系数。

$$r_{xy}=\frac{\operatorname{cov}(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{8}$$

$$\begin{aligned}&\operatorname{cov}(x,y)\\&=\frac{1}{N}\sum_{i=1}^{N}(x_i-E(x))(y_i-E(y)),D(x)\\&=\frac{1}{N}\sum_{i=1}^{N}(x_i-E(x))^2,E(x)=\frac{1}{N}\sum_{i=1}^{N}x_i\end{aligned} \tag{9}$$



(a1) lena_gray (a2) 水平方向相关性 (a3) 垂直方向相关性 (a4) 对角线方向相关性

(a) lena_gray明文图像相关性

(b1) lena_gray密文图像 (b2) 水平方向相关性 (b3) 垂直方向相关性 (b4) 对角线方向相关性

(b) lena_gray密文图像相关性

图 3 lena_gray 相关性

Figure 3 Correlation of lena_gray

计算结果如表 3 所示。从表中可知明文图像的相关性系数接近 1, 而密文图像的相关性系数接近 0。表 4 是所提出的算法与近期代表性图像加密算法的相关性比较, 可以看到与其他方法相比, 本文算法相关性系数相对较为接近 0。这说明我们的算法具有抵抗相关性分析的能力。

4.4 信息熵分析

信息熵可表示图像中像素值分布的随机性。本文信息熵还可以表示 DNA 序列中 DNA 码分布的随机性。我们用式(10)计算信息熵。式中 a_i 表示像素或 DNA 编码, $p(a_i)$ 表示 a_i 的概率值, l 表示 255 或 63。

$$H(a)=\sum_{i=0}^{l}p(a_i)\log_2\frac{1}{p(a_i)} \tag{10}$$

表 5 是图像和 DNA 序列的信息熵, 表 6 是本文算法与其他方法的信息熵比较。从表中我们可以看到, 密文图像的信息熵趋近 8, 密文 DNA 序列的信

表 3 相关性系数

Table 3 Correlation coefficients

	明文图像			密文图像		
	水平	垂直	对角线	水平	垂直	对角线
lena_gray	0.9502	0.9849	0.9943	–0.0005	–0.0077	–0.0004
peper_gray	0.9843	0.9841	0.9992	0.0016	0.0032	–0.0007

表 4 相关性系数比较

Table 4 Comparison of correlation coefficients

图像	lena_gray	文献[14]	文献[15]	文献[16]
水平	–0.0005	0.0022	0.0023	–0.0070
垂直	–0.0077	0.0009	–0.0020	0.0083
对角线	–0.0004	0.0012	–0.0073	0.0013

表 5 信息熵

Table 5 Information entropy

	图像		DNA 序列	
	lena_gray	peper_gray	lena_gray	peper_gray
明文	7.3871	7.1849	5.9246	5.9294
密文	7.9983	7.9983	5.9999	5.9999

表 6 密文图像信息熵比较

Table 6 Comparison of information entropy

图像	lena_gray	文献[14]	文献[15]	文献[21]
信息熵	7.9983	7.9971	7.9960	7.9376

息熵趋近 6(每个 DNA 编码对应图像的 6 比特信息)。与其他方法相比, 所提算法密文图像信息熵较为较为接近理想值 8。因此在加密后, 密文图像和密文序列的信息分布具有良好的随机性。

4.5 抗差分攻击分析

差分攻击指攻击者向加密器输入具有微小差别的一系列明文来观察得到的密文, 以反推加密器结构的攻击方式。衡量图像加密系统是否能抵御差分攻击的指标是平均像素改变率(NPCR)和平均像素改变强度(UACI)。本文 NPCR, UACI 也可分别表示平均 DNA 码改变率和平均 DNA 码改变强度。计算过程中, 我们按照 3.5 节中编码和数字的映射关系来量化 DNA 码。

本文密钥 x_0, y_0, z_0 由明文图像的哈希值产生, 所以本文用密钥的微小改变来代替明文的微小变化。我们用(11)式和(12)式来计算 NPCR 和 UACI。式中 $m(i)$, $m(i')$ 表示两个不同密文在 i 位置的像素/DNA 码。如果 $m(i)$ 等于 $m(i')$, 则 $E(m(i),m(i'))=0$, 反之, $E(m(i),m(i'))=1$。S 表示像素点或 DNA 码总数。l 表示 255 或 63。

$$\mathrm{NPCR}=\frac{\sum_{i,i'}E(m(i),m(i'))}{S}\times 100\% \tag{11}$$

$$\mathrm{UACI}=\frac{1}{S}\left[\sum_{i,i'}\frac{|m(i),m(i')|}{l}\right]\times 100\% \tag{12}$$

表 7 密文 NPCR 和 UACI

Table 7 NPCR and UACI of ciphertexts

	lena 密文图像		peper 密文图像		lena 密文序列		peper 密文序列	
	NPCR(%)	UACI(%)	NPCR(%)	UACI(%)	NPCR(%)	UACI(%)	NPCR(%)	UACI(%)
(x_0+10^{-14}, y_0, z_0)	99.61	29.19	99.62	29.32	98.43	33.27	98.44	33.35
(x_0, y_0+10^{-14}, z_0)	99.61	29.12	99.61	29.31	98.43	33.33	98.44	33.41
(x_0, y_0, z_0+10^{-14})	99.59	29.14	99.62	29.32	98.44	33.29	98.44	33.27

表 7 为密文图像与密文序列的 NPCR 和 UACI。从表中可以看到, 密文图像 NPCR 接近期望值 100%, 密文序列 UACI 接近期望值 33.46%。密文图像的 UACI 小于理想值是因为在扩散中我们以 6 比特为单位进行扩散, 而像素包含 8 比特。密文序列 NPCR 小于期望值是因为我们只有 64 个编码, 变化空间较小。

文献[21]同样是面向 DNA 存储的加密算法, 文中给出 lena 密文图像的 NPCR 和 UACI 分别为 29.15%和 1.37%。与文献[21]相比, 本文的方案抗差分攻击能力更强。

4.6 经典攻击类型分析

经典攻击类型包括唯密文攻击、已知明文攻击、选择密文攻击和选择明文攻击。算法如能抵抗选择明文攻击, 就必能抵抗其他三种攻击。本文的密钥由明文图像得到, 可近似达到一次一密效果。因此算法可抵抗选择明文攻击以及其它三种经典攻击。

4.7 鲁棒性分析

传统图像加密算法常以抗剪辑及抗噪声能力来分析算法的鲁棒性。本文为面向 DNA 存储的加密算法。DNA 存储中序列的缺失和碱基的插入、删除、

替换错误是现阶段存储系统无法避免的问题。所以本文的鲁棒性分析从碱基错误和序列缺失两个方面讨论。

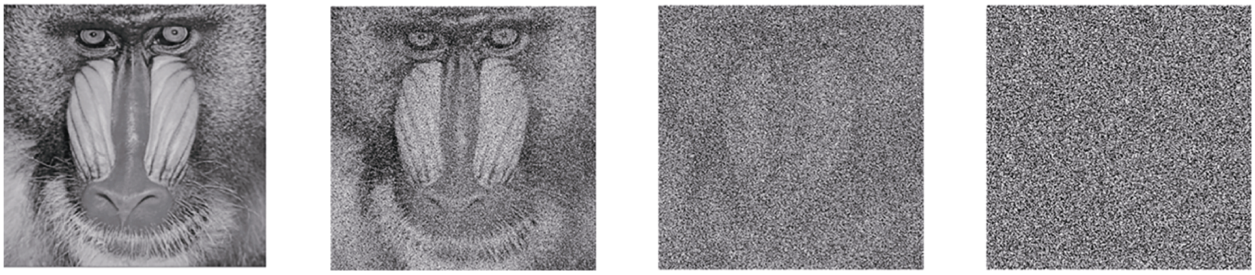
4.7.1 抗碱基错误能力分析

图 4(a)为密文序列经过纠错后解密得到的图像, 其中碱基错误包括插入、删除和替换。从图 4(a)可以看到, 错误率为 10%时, 解密图像十分清晰。错误率为 20%时, 解密图像仍然能够辨识。第三代 DNA 测序技术错误率为 10%～15%, 因此该方案在现有的 DNA 存储系统中具有较强的鲁棒性。

图 4(b)为密文序列未经过纠错而解密得到的图像。图 4 (b2)和图 4 (b3)错误类型只有替换。图 6(b4)包含三种错误类型。从图 4 (b2) 、图 4 (b3)中可以看到, 1%的替换错误不影响图像的识别, 而 10%的替换错误会导致图像无法识别。从图 4 (b4)可以看到, 错误率为 0.1%时, 解密图像无法识别。上述现象是因为替换错误只影响序列中的单个 DNA 码, 而插入和删除错误则会影响后续所有 DNA 码。所以当错误率较低且只存在替换错误时, 不影响图像识别; 当存在插、删错误时, 即使错误率很低, 图像也无法识别。

(a1) baboon_gray (a2) 1%插、删、替错误 (a3) 10%插、删、替错误 (a4) 20%插、删、替错误

(a) 纠错后解密图像



(b1) baboon_gray (b2) 1%替换错误 (b3) 10%替换错误 (b4) 0.1%插、删、替错误

(b) 未纠错解密图像

图 4 baboon_gray 碱基错误分析

Figure 4 Analysis of base error on baboon_gray

由此我们可以得到一个结论: 面向 DNA 存储的加密算法, 如果没有纠错能力将会使解密变得十分困难。

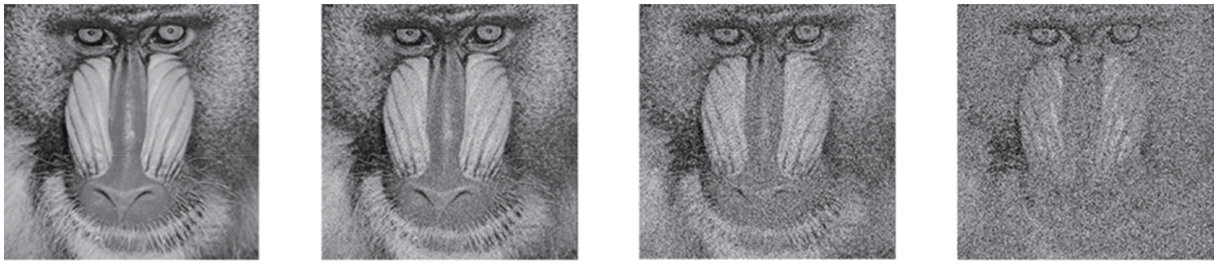
4.7.2 抗序列缺失能力分析

图 5 为缺失一部分序列解密得到图像。从图 5(b)和图 5(c)中可以看到, 密文序列缺失 1/16 或 1/4 时, 图像较为清晰。从图 5(d)中可以看到, 序列缺失 1/2 时, 解密图像仍具有一定辨识度。因此本文的算法具有较强的抗序列缺失能力。

值得注意的是, 本算法所使用的纠错方案不能处理序列缺失问题。抗序列缺失能力来自于加密算法整体的鲁棒性。这说明本文的置乱方案以及扩散方案具有较强的鲁棒性, 可以抵抗 DNA 存储环境中的序列缺失问题。

5 总结

随着 DNA 存储技术的快速发展, 融合分子生物学与现代密码学的 DNA 存储加密技术也逐渐引起人们的重视。然而, 目前图像 DNA 存储加密研究较少考虑 DNA 存储环境特有复杂错误带来的影响, 阻碍了这些方案的有效实施。本文提出了一种基于前向纠错码的图像 DNA 加密存储算法。所提算法主要包括两个部分: (1) 利用混沌系统、像素点置乱、图像分解等现代加密技术, 在二进制水平上对图像像素进行置乱和扩散; (2) 动态使用基于前向纠错码的 DNA 编码表对二进制密文进行 DNA 碱基映射, 进一步扩大密钥空间, 并对 DNA 存储过程中的错误具有一定的容忍力。仿真结果表明, 所提算法可以有效抵



(a) baboon_gray (b) 密文序列缺失1/16 (c) 密文序列缺失1/4 (d) 密文序列缺失1/2

图 5 baboon_gray 抗密文序列缺失分析

Figure 5 Analysis of DNA sequence loss on baboon_gray

抗差分攻击、已知明文攻击等多种密码学攻击; 对 DNA 存储过程中的错误具有较强的鲁棒性, 在碱基错误率 20%或序列缺失比例 50%的条件下解密图像依然清晰可见。下一步我们将深入研究可用于加密存储的生物分子特殊作用机制, 探索出一个适用于 DNA 存储环境、集现代密码学与分子生物学于一体的专用存储加密方法, 解决通用文件在 DNA 存储环境的安全传输问题。

参考文献

[1] Gao Y M, Chen X, Qiao H Y, et al. Low-Bias Manipulation of DNA Oligo Pool for Robust Data Storage[J]. *ACS Synthetic Biology*, 2020, 9(12): 3344-3352.

[2] Heckel R, Mikutis G, Grass R N. A Characterization of the DNA Data Storage Channel[J]. *Scientific Reports*, 2019, 9: 9663.

[3] Takahashi C N, Nguyen B H, Strauss K, et al. Demonstration of End-to-End Automation of DNA Data Storage[J]. *Scientific Reports*, 2019, 9: 4998.

[4] Blawat M, Gaedke K, Hütter I, et al. Forward Error Correction for DNA Data Storage[J]. *Procedia Computer Science*, 2016, 80: 1011-1022.

[5] Organick L, Ang S D, Chen Y J, et al. Random Access in Large-Scale DNA Data Storage[J]. *Nature Biotechnology*, 2018, 36(3): 242-248.

[6] Xu P, Fang G, Shi X L, et al. DNA Storage and Its Research Progress[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1326-1331.
(许鹏, 方刚, 石晓龙, 等. DNA 存储及其研究进展[J]. *电子与信息学报*, 2020, 42(6): 1326-1331.)

[7] Huo J J, Zhang W Z. DNA Cryptography and Application of DNA Computing[J]. *Journal of China Academy of Electronics and Information Technology*, 2014, 9(1): 17-21.
(霍家佳, 张文政. DNA 密码与 DNA 计算及应用[J]. *中国电子科学研究院学报*, 2014, 9(1): 17-21.)

[8] Yao X Y, Zan X Z, Xie L, et al. An Overview of the Complexity of DNA Storage[J]. *Journal of Guangzhou University* (*Natural Science Edition*), 2021, 20(1): 12-22.
(姚翔宇, 昝乡镇, 谢恋, 等. DNA 存储技术的复杂度概述[J]. *广州大学学报(自然科学版)*, 2021, 20(1): 12-22.)

[9] Zan X Z, Yao X Y, Xu P, et al. A Survey on Error Correcting Algorithms in DNA Storage[J]. *Journal of Guangzhou University* (*Natural Science Edition*), 2021, 20(2): 13-22.
(昝乡镇, 姚翔宇, 许鹏, 等. DNA 存储中的纠错方法综述[J]. *广州大学学报(自然科学版)*, 2021, 20(2): 13-22.)

[10] Clelland C T, Risca V, Bancroft C. Hiding Messages in DNA Microdots[J]. *Nature*, 1999, 399(6736): 533-534.

[11] Gehani A, LaBean T, Reif J. DNA-Based Cryptography[M]. Aspects of Molecular Computing. Berlin, Heidelberg: Springer, 2003: 167-188.

[12] Lu M X, Lai X J, Xiao G Z, et al. Symmetric Encryption Method Based on DNA Technology[J]. *Science in China* (*Series E* (*Information Sciences*)), 2007, 37(2): 175-182.
(卢明欣, 来学嘉, 肖国镇, 等. 基于 DNA 技术的对称加密方法[J]. *中国科学(E 辑: 信息科学)*, 2007, 37(2): 175-182.)

[13] Lai X J, Lu M X, Qin L, et al. Asymmetric Encryption and Signature Method Based on DNA Technology[J]. *Scientia Sinica* (*Informations*), 2010, 40(2): 240-248.

[14] Lai Xuejia, Lu Mingxin, Xiao Guozhen, et al. Asymmetric encryption and signature method based on DNA technology[J]. *Science in China Series F-Information Sciences*, 2010, 40(2): 240-248.
(来学嘉, 卢明欣, 秦磊, 等. 基于 DNA 技术的非对称加密与签名方法[J]. *中国科学*, 2010, 40(2): 240-248.)

[15] Wang X Y, Zhang Y Q, Zhao Y Y. A Novel Image Encryption Scheme Based on 2-D Logistic Map and DNA Sequence Operations[J]. *Nonlinear Dynamics*, 2015, 82(3): 1269-1280.

[16] Wang X Y, Su Y N. Image Encryption Based on Compressed Sensing and DNA Encoding[J]. *Signal Processing*: *Image Communication*, 2021, 95: 116246.

[17] Thangavel M, Varalakshmi P. Enhanced DNA and ElGamal Cryptosystem for Secure Data Storage and Retrieval in Cloud[J]. *Cluster Computing*, 2018, 21(2): 1411-1437.

[18] Niu Y, Zhang X C. Image Encryption Algorithm of Based on Variable Step Length Josephus Traversing and DNA Dynamic Coding[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1383-1391.
(牛莹, 张勋才. 基于变步长约瑟夫遍历和 DNA 动态编码的图像加密算法[J]. *电子与信息学报*, 2020, 42(6): 1383-1391.)

[19] Zakeri B, Carr P A, Lu T K. Multiplexed Sequence Encoding: A Framework for DNA Communication[J]. *PLoS ONE*, 2016, 11(4): e0152774.

[20] Zhang Y N, Wang F, Chao J E, et al. DNA Origami Cryptography for Secure Communication[J]. *Nature Communications*, 2019, 10: 5469.

[21] Castro C E, Kilchherr F, Kim D N, et al. A Primer to Scaffolded DNA Origami[J]. *Nature Methods*, 2011, 8(3): 221-229.

[22] Peng W P, Cui S A, Song C. One-Time-Pad Cipher Algorithm Based on Confusion Mapping and DNA Storage Technology[J]. *PLoS ONE*, 2021, 16(1): e0245506.

[23] Rossler O E. An Equation for Hyperchaos[J]. *Physics Letters A*, 1979, 71(2/3): 155-157.

[24] Bertoni G, Daemen J, Peeters M, et al. Keccak[C]. *Annual international conference on the theory and applications of cryptographic techniques*, 2013: 313-314.

[25] Zan X Z, Yao X Y, Xu P, et al. A Hierarchical Error Correction Strategy for Text DNA Storage[J]. *Interdisciplinary Sciences: Computational Life Sciences*, 2022, 14(1): 141-150.

[26] Zan X Z, Yao X Y, Xu P, et al. An Efficient Bueket-Allocation Decoding Method Based on Forward Error Correction Codes for Deoxyribo Nucleicecid Storage[J]. *Journal of Electronics & Information Technology*, 2022, 44(10): 3650-3656.
(昝乡镇, 姚翔宇, 许鹏, 等. 一种高效的前向纠错码桶分配 DNA 存储解码方法[J]. *电子与信息学报*, 2022, 44(10): 3650-3656.)

姚翔宇 于 2018 年在大连理工大学大学计算机科学与技术专业获得工学学士学位。现在广州大学网络空间安全专业攻读硕士学位。研究领域为网络安全、DNA 存储。研究兴趣包括: 机器学习、密码学。Email: 1746547770@qq.com

苏燕青 于 2020 年在国际关系学院网络空间安全学院获得工学学士学位。现在广州大学计算机科学与网络工程学院攻读工学硕士学位。研究领域为 DNA 存储、信息安全等。Email: 2112106051@e.gzhu.edu.cn

昝乡镇 现在广州大学网络空间安全专业攻读博士学位。研究领域为 DNA 存储、信息安全。Email: xiangcheng2436@163.com

许鹏 于 2018 年在东南大学生物医学工程专业获得工学博士学位。现任广州大学计算科技研究院副教授。研究兴趣为生物信息、机器学习、DNA 存储等。Email: gdxupeng@gzhu.edu.cn

刘文斌 现任广州大学计算机科学与网络工程研究院教授。主要研究方向: DNA 存储, 信息安全。Email: wbliu6910@gzhu.edu.cn