基于混沌映射与有限域 GF(2^4)域乘法运算的电子病历图像的加密

刘西林 严广乐
(上海理工大学管理学院 上海 200093)

摘 要 针对电子病历的保密性问题,提出一种混沌映射与 GF(2^4)域乘法运算相结合的电子病历图像加密算法。通过对电子病历的灰度图像使用 SHA-1 算法产生的哈希值作为病历摘要来监测电子病历的传播。利用二维图像展成一维向量后的无重复置乱算法结合 GF(2^4)域乘法运算的扩散算法对病历图像进行加密。Lorenz 混沌映射产生相应的密码。实验结果表明:算法的安全性高,有效保证电子病历在传递过程中的安全性。

关键词 电子病历 混沌映射 SHA-1 病历摘要 GF(2^4)

中图分类号 TP309.7 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2018.12.056

ENCRYPTION OF ELECTRONIC MEDICAL RECORD IMAGE BASED ON CHAOTIC MAPPING AND MULTIPLICATION OF GF (2^4) DOMAIN

Liu Xilin Yan Guangle
(*Business School*, *University of Shanghai for Science and Technology*, *Shanghai 200093*, *China*)

Abstract For the confidentiality problem of electronic medical records, we proposed the encryption algorithm of electronic medical records image based on chaotic mapping and multiplication of GF(2^4) domain. For grayscale images of electronic medical records, we took the hash value generated by SHA-1 algorithm as the medical record summary to monitor the spread of electronic medical record. After the two-dimensional image was expanded into one-dimensional vector, the non-repetitive scrambling algorithm was combined with the diffusion algorithm of multiplication algorithm in GF(2^4) domain to encrypt the medical record image. Lorenz chaotic map generated the corresponding cryptogram. The results show that the algorithm has high security and effectively ensures the safety of electronic medical records in the process of transition.

Keywords Electronic medical record Chaotic mapping SHA-1 Medical record summary GF(2^4)

0 引 言

随着互联网 + 技术的广泛应用和医院信息化的不断发展,电子病历已成为医院信息化进程中的必然结果,保证病历信息的真实性、完整性和对患者隐私信息的保护已成为当前电子病历中的热点问题[1]。随着健康信息技术的发展,医院之间可能就相关医患信息进行传递、交流,减少患者重复检验、检查及用药,提升医疗资源运用效率[2]。这在另一方面也增加了电子病历在操作、存储和传递方面的风险,使得电子病历越来越容易被盗取、复制、外泄、篡改。因此,对电子病历加密的研究具有重大意义。

由于图像具有冗余度高、数据量大、像素间相关性强等特点,传统的加密算法不太适用于图像加密,图像加密需要使用快速的方法[3]。为了提高数字图像的安全度,近年来对数字图像加密的研究很多[4],但有的存在密钥空间小的问题,有的存在信息熵偏差大的问题。电子病历图像作为携带病人信息的图像,目前对它的加密研究很少,而周广彬等[7]提出的混沌加密电子病历虽然达到了加密电子病历图像的效果,但是仅使用了 Henon 混沌算法进行加密。Henon 作为简单的二维非线性混沌系统,有着低维混沌系统密钥空间小、安全性不高的缺点[8]。而且周广彬等对电子病历信息使用

收稿日期:2018 - 06 - 19。上海高原学科建设项目管理科学与工程(10 - 17 - 303 - 004)。**刘西林**,硕士生,主研领域:混沌加密,图像处理。**严广乐**,教授。

的是 MD5 算法产生的病历摘要,但是 MD5 算法已经被破解,从而使得安全性存在问题。

本文通过对电子病历图像使用 SHA-1 算法生成 160 bit 的哈希值作为病历摘要来监测电子病历传播的安全性;使用二维图像展成一维向量的无重复置乱和 GF(2^4)域乘法的两次不同的扩散算法对电子病历图像进行加密,而后生成密文图像;三维 Lorenz 混沌映射产生密码。传输密文图像就达到了传播电子病历的目的,这样可以有效防止电子病历外泄,实现了对电子病历的隐秘传输[9]。这也使得不法分子看到的是密文图像,并不能看到真实的电子病历图,另一方面解密后的秘密图像再次经过 SHA-1 算法编码产生的病历摘要和发送方的病历摘要匹配,可以检查电子病历是否在传播过程中被私自篡改。这些方法使电子病历在传播中得到了多重保护。

1 SHA-1 算法、混沌映射与 GF(2^4)算法

1.1 SHA-1 算法

SHA 算法是密码散列函数家族,是经过 FIPS 所认证的安全散列算法。SHA-1 算法就是其中一个,它可以将明文信息转换成字符串,再进行补位操作,然后附加长度通过函数计算得出唯一的 160 bit 的信息摘要[10]。图像的明文信息出现任何微小的变化经过 SHA-1 编译的哈希值即病历摘要都会发生显著的变化。每张电子病历经过使用 SHA-1 算法相当于拥有了“指纹”。

1.2 混沌 Lorenz 系统

本文采用的是 Lorenz 系统映射。其具体的动力学方程如下所示:

$$\begin{cases} x = a(y - x) + w \\ y = cx - y - xz \\ z = xy - bz \\ w = -yz + rw \end{cases} \tag{1}$$

式中:a、b、c、w、r 为混沌系统的参数,当 $a = 10$,$b = 8/3$,$c = 28$,$-1.52 \leqslant r \leqslant -0.06$ 时,式(1)处于混沌状态[11]。

1.3 GF(2^4)算法

在密码学中,GF(p)即伽罗华域,是一个非常重要的有限域,并且域中必须有单元。GF(p)即 mod p,p 为素数,结果是有限域中元素。在实际应用中,为了防止数据丢失,引入了 GF(p^w),其中 p 为素数,通常为 2。伽罗华域的元素可以通过该域上的本原多项式生成,通过本原多项式得到的域,其加法单元是 0,乘法单元是 1。在 GF(2^4)域中取既约多项式 $m(x) = x^4 + x + 1$。

2 加密与隐藏方案设计

方案设计流程图如图 1 所示。

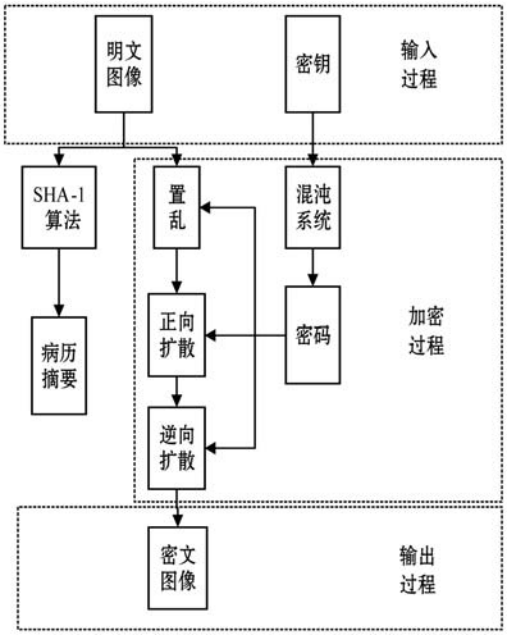


图 1 方案设计图

2.1 产生病历摘要

将电子病历图像转换成灰度图像,再经过 SHA-1 算法产生的哈希值作为病历摘要,值为 42817e38ab192e6b3bb2491578ab3cf65a5cf7ec。发送方保存该病历摘要,随后与接收方解密电子病历图像再经过 SHA-1 算法编码产生的哈希值进行匹配。

2.2 电子病历图像加密

本文提出的加密算法首先对原始图像的像素点的位置进行置乱操作;然后改变图像的灰度值进行正向扩散与逆向扩散;混沌 Lorenz 系统的参数和初始值作为密钥,产生对应的密码。解密过程是加密过程的逆过程。假设原始图像矩阵大小为 $M \times N$,具体加密步骤如下:

1) 给定密钥 K 的值即混沌 Lorenz 系统的各个变量的值(初值 x_0、y_0以及 z_0和参数值 w_0)。迭代超混沌系统产生长度为 $M \times N$ 浮点数形式的伪随机序列。

2) 将明文图像矩阵按行展开成一维向量,记作 $\boldsymbol{A}$。借助于混沌系统产生的 $M \times N$ 的伪随机序列 x_i,$i = 1,2,\cdots,M \times N$,$X$ 中重复出现的伪随机数只保留第一个,将$\{1,2,\cdots,M \times N\}$中没有 X 中的数值按从小到大的顺序添加到 X 的末尾,最后交换 $A(x_i)$ 与 $A(x_{MN-i+1})$的位置,从而完成了二维图像展成一维向

量的无重复置乱算法。

3）置乱算法之后对像素点的灰度值采用 GF(2^4)域乘法运算的扩散算法,本文采用的是正向扩散与逆向扩散相结合的方法。正向扩散和逆向扩散如下所示:

$$\begin{cases} \boldsymbol{C}_{i,H} = \boldsymbol{C}_{i-1,H} \times \boldsymbol{S}_{i,H} \times \boldsymbol{P}_{i,H} \\ \boldsymbol{C}_{i,L} = \boldsymbol{C}_{i-1,L} \times \boldsymbol{S}_{i,L} \times \boldsymbol{P}_{i,L} \\ \boldsymbol{C}_i = (\boldsymbol{C}_{i,H} \times 16 + \boldsymbol{C}_{i,L}) \end{cases} \tag{2}$$

$$\begin{cases} \boldsymbol{C}_{i,H} = \boldsymbol{C}_{i+1,H} \times \boldsymbol{S}_{i,H} \times \boldsymbol{P}_{i,H} \\ \boldsymbol{C}_{i,L} = \boldsymbol{C}_{i+1,L} \times \boldsymbol{S}_{i,L} \times \boldsymbol{P}_{i,L} \\ \boldsymbol{C}_i = (\boldsymbol{C}_{i,H} \times 16 + \boldsymbol{C}_{i,L}) \end{cases} \tag{3}$$

式中:$\boldsymbol{P}$ 为明文展开成的一维向量,$\boldsymbol{C}$、$\boldsymbol{S}$ 为密码向量,初值 C_0、S_1来自于密钥,$i = 1,2,\cdots,MN$。H 表示数据的高 4 位,L 表示数据的低 4 位。

经过一次置乱和正向与逆向两次不一样的扩散,从而得到了电子病历图像的密文图像。解密是加密的逆过程,不再赘述。

3　仿真实验

在 MATLAB 7.1 环境下对本文提出的算法进行仿真实验,得出结果。原始电子病历明文图像如图 2 所示,加密后电子病历密文图像如图 3 所示,正确密钥解密密文图像如图 4 所示,错误密钥解密密文图像如图 5 所示。

图 2　电子病历明文图像

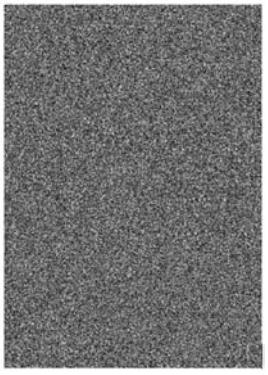


图 3　电子病历密文图像

图 4　正确密钥解密密文图像

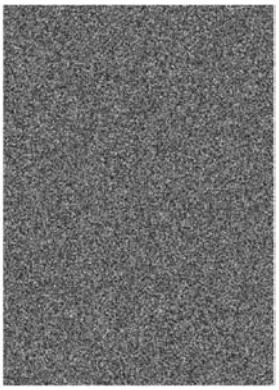


图 5　错误密钥解密密文图像

接收方收到电子病历图像密文经过解密操作,再使用 SHA-1 算法编码产生的哈希值为:42817e38ab192e6b3bb2491578ab3cf65a5cf7ec。若在传播中图像有改动再次编码得出的哈希值为:b3df52208a35ba5a0ada56862a07a7a0b7f9d3bb(改动位置不同,哈希值不同)。

接收方将得到的哈希值与发送方的病历摘要进行匹配,若存在差异,说明该病历图像在传播中存在被篡改的行为;若相同,则说明病历图像安全传输。

4　安全性分析

4.1　直方图与 x^2检验分析

加密可以将明文图像转换成噪声从而隐藏信息。直方图与直方图的 x^2检验可以描述图像的相关性。一般情况下,图像像素灰度的直方图越服从均匀分布,x^2检验值越小,越能有效地抵抗统计分析的攻击。电子病历图像明文直方图如图 6 所示,电子病历图像密文直方图如图 7 所示,电子病历图像明文与密文的 x^2值如表 1 所示。从图与表可知密文图像的像素灰度值更接近于均匀分布,说明加密效果比较好。

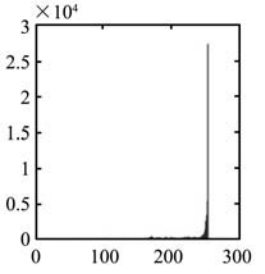


图 6　明文直方图

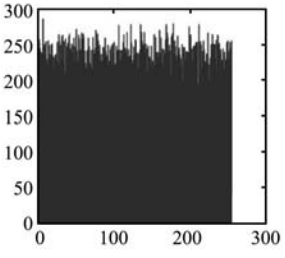


图 7　密文直方图

表 1　x^2检验结果

图像	x^2值
明文	3.244 2e +06
密文	267.321 9

4.2　相关性分析

明文图像相邻像素之间有很强的相关性,而这些相关性内部存在着明文的部分信息,若被不法分子发现利用,很可能会造成图像的泄露[12]。好的加密算法能够使得图像的像素之间的相关性变弱。本文从明文与密文图像中随机挑选了 2 000 对相邻像素点,绘画出相关性图像如图 8 所示,计算出了它们在水平、垂直、正对角与反对角的相关系数如表 2 所示。从图与表可以看出加密后图像像素之间的相关性明显降低,有效地保护了图像信息。相关系数的计算公式为:

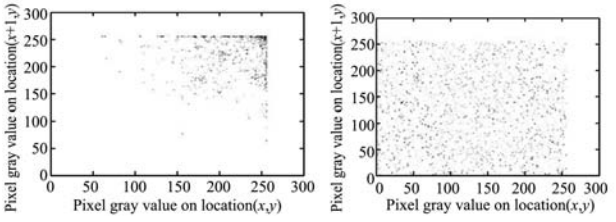
$$r_{xy} = \frac{\operatorname{cov}(\boldsymbol{u},\boldsymbol{v})}{\sqrt{D(\boldsymbol{u})}\sqrt{D(\boldsymbol{v})}} \tag{4}$$

其中，$\mathrm{cov}(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(\boldsymbol{u}))(y_i - E(\boldsymbol{v}))$

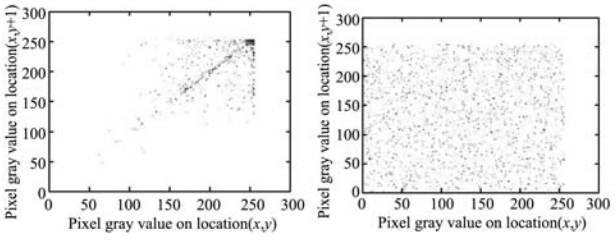
$$D(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N}(u_i - E(\boldsymbol{u}))^2$$

$$E(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N}u_i$$

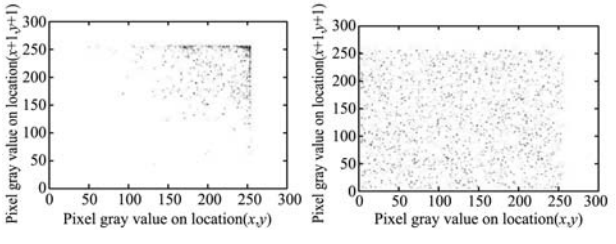
式中：N 为任取的相邻像素点的对数，它们的灰度值为 (u_i,v_i)，$i=1,2,\cdots,N$，向量 $\boldsymbol{u}=\{u_i\}$，向量 $\boldsymbol{v}=\{v_i\}$。



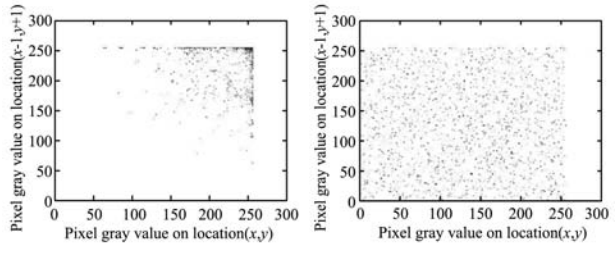
明文水平方向(a)　密文水平方向(b)



明文垂直方向(c)　密文垂直方向(d)



明文的正对角方向(e)　密文的正对角方向(f)



明文的反对角方向(g)　密文的反对角方向(h)

图 8　相关性图像

表 2　相关系数

图像	水平	垂直	正对角	反对角
明文	0.274 1	0.775 5	0.117 9	0.142 5
密文	0.015 2	−0.028 6	0.003 8	0.007 1

4.3　信息熵计算

信息熵反映的为图像信息的不确定性，一般认为，图像信息熵越大，信息量越大，事件的随机性越大[11]。为了体现本文算法的优越性，不仅计算出电子病历图像的明文与密文的信息熵，而且用本文算法加密经典图像 Lena 与其他文献进行对比，具体的结果如表 3 所示。信息熵的计算公式为：

$$H = -\sum_{i=0}^{L}P(i)\log 2P(i) \tag{5}$$

式中：L 为图像灰度等级数，$P(i)$ 表示灰度值 i 出现的概率。

对于 $L=256$ 的灰度图像，信息熵 H 理论值为 8，因为仿真实验得到电子病历的图像密文的信息熵几乎为 8。又经过本文算法、文献[4]算法和文献[6]算法加密同一幅图像 Lena 计算信息熵，对比可以看出本算法的信息熵更接近 8，表示本加密算法更能有效地抵抗数据攻击[13]。

表 3　信息熵结果

图像	信息熵值
电子病历图像明文	4.288 0
电子病历图像密文	7.997 5
Lena 图像明文	7.380 2
本算法加密 Lena 图像密文	7.999 4
文献[4]加密 Lena 图像密文	7.999 2
文献[6]加密 Lena 图像密文	7.997 0

4.4　密钥空间分析

密钥空间是指所有合法密钥的集合，加密算法越好，密钥空间越大[14]。本文的密钥为 Lorenz 系统的初始值，即 $K=\{x_0,y_0,z_0,w_0\}$，其中 $x_0\in(-40,40)$、$y_0\in(-40,40)$、$z_0\in(1,81)$、$w_0\in(-250,250)$，x_0、y_0 和 z_0 的步长为 10^{-13}，w_0 的步长为 10^{-12}，可得密钥空间大约为 2.56×10^{59}，密钥空间约为 197 bit，而文献[5]的密钥空间大小为 $(10^{16})^2$，因此，本算法的密钥空间更大，抵抗暴力攻击更有效。

4.5　密钥敏感性分析

密钥敏感性分析旨在将密钥做微小变化后，再加密同一图像得到的密文图像，若密文图像存在显著差别，则称密钥敏感性强，反之，密钥敏感性则弱[11]。衡量大小相同图像差别有几个常用指标：NPCR 记录不同的像素点个数占全部像素点的比例，具体公式如下：

$$NPCR(P_1,P_2) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|Sign(P_1(i,j) - P_2(i,j))|\times 100\% \tag{6}$$

式中：$Sign(x)=\begin{cases}1 & x>0\\0 & x=0\\-1 & x<0\end{cases}$

UACI 记录两幅图像相应像素点的差值与最大差

值(255)比值的平均值,具体公式如下:

$$UACI(P_1,P_2) = \frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\frac{|P_1(i,j) - P_2(i,j)|}{255 - 0} \times 100\% \quad (7)$$

BACI 首先求得两幅图像的差图像的绝对值,然后将图像分解,计算全部小图像任意两个像素点的差值的绝对值的平均值与像素最大差值(255)的比值,具体公式如下(假设图像大小为 $M \times N$ 的 $P1$ 和 $P2$ 两幅图像):

$$BACI(P_1,P_2) = \frac{1}{(M-1)(N-1)}\sum_{i=1}^{(M-1)(N-1)}\frac{m_i}{255} \quad (8)$$

式中:m 为小图像块,$i = 1,2,\cdots,(M-1)(N-1)$。

本文从密钥空间中随机选取 1 000 个值,分别对 x_0、y_0与 z_0改变 10^{-13},w_0改变 10^{-12},计算的 1 000 个 NPCR、UACI 与 BACI 的平均值如表4 所示。本文计算指标的结果很接近它们的理论期望值 99.609 4%、33.463 5%与 26.771 2%,说明密钥发生微小的变化后,密文相差很大,也进一步说明本加密算法密钥敏感性强,具有很强的抗差分能力。

表 4　密钥敏感性分析结果　　%

初值	指标		
	NPCR	UACI	BACI
x_0	99.509 0	33.428 3	26.745 7
y_0	99.609 0	33.466 7	26.729 0
z_0	99.609 3	33.464 4	26.772 1
w_0	99.609 5	33.467 0	26.776 7
理论值	99.609 4	33.463 5	26.771 2

5　结　语

本文针对电子病历传输中的安全性问题提出了基于混沌映射与 GF(2^4)域乘法运算的电子病历图像的加密算法。该算法通过对传输的电子病历图像进行加密与监测相结合的方法,增强了传输中电子病历图像的安全性与可靠性。实验结果表明,本算法密钥空间大,能有效抵抗暴力、统计以及差分攻击。该算法以后会有很强的潜在应用价值,但在提高加密与解密效率的问题上需要进一步提高,这个问题是今后需要研究的方向。

参　考　文　献

[1] 王其琳,侯冷晨,郑军华,等. 调查分析医患双方对电子病历应用的认知[J]. 现代医院管理,2015(5):72-74.

[2] 武萌. 电子病历与患者隐私权保护[J]. 继续医学教育,2017,31(7):85-87.

[3] 朱淑芹,李俊青,王文宏. 对改进的基于 DNA 编码和混沌的图像加密算法的安全性分析[J]. 计算机应用研究,2017,34(10):3090-3093.

[4] 闫兵,柏森,刘博文,等. 基于交叉混沌映射的小波域图像加密算法[J]. 计算机应用研究,2018,35(6): 1797-1799,1811.

[5] 李春虎, 罗光春, 李春豹. 基于斜帐篷混沌映射和 Arnold 变换的图像加密方案[J/OL]. 2018, 35(11). [2017-11-10]. http://www.arocmag.com/article/02-2018-11-028.html.

[6] Zhang M, Tong X. A new algorithm of image compression and encryption based on spatiotemporal cross chaotic system [J]. Multimedia Tools & Applications, 2015, 74(24): 11255-11279.

[7] 周广彬,姚磊,陈晓军,等. 混沌加密和隐藏在电子病历中的引用[J]. 电脑知识与技术,2017,13(20):176-177.

[8] 郭凤鸣,涂立. 混沌理论在密码学中的应用[M]. 北京. 北京理工大学出版社,2015:26.

[9] 周思成,翟晓梅. 电子医疗保健情境下的隐私保护[J]. 中国医学伦理学,2016,29(4):681-684.

[10] 刘坤,杨正校. 基于局部碰撞算法的 SHA-1 改进算法设计与研究[J]. 软件工程,2017,20(11):27-29.

[11] 张勇. 混沌数字图像加密[M]. 北京. 清华大学出版社,2016:16-17.

[12] 王宏达. 一种基于混沌系统的新型图像加密算法[J]. 光学技术,2017,43(3):260-266.

[13] Zhu H, Zhao C, Zhang X. A novel image encryption-compression scheme using hyper-chaos and Chinese remainder theorem [J]. Signal Processing: Image Communication, 2013,28(6):670-680.

[14] 朱淑芹, 李俊青, 葛广英. 基于一个新的五维离散混沌的快速图像加密算法[J]. 计算机科学, 2016, 43(S2): 411-416.

(上接第 192 页)

[15] Suzuki J, Nagata M. Cutting-off redundant repeating generations for neural abstractive summarization[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017.

[16] Suzuki J, Nagata M. RNN-based encoder-decoder approach with word frequency estimation[EB]. eprint arXiv:1701.00138, 2016.

[17] Lin C Y. ROUGE: Recall-oriented understudy for gisting evaluation[J]. 2003.