研究与开发

# 基于动态有限域的保留格式加密算法

元河清[1]，申锦尚[1]，王强[2]
（1. 山东新潮信息技术有限公司，山东 济南 250100；
2. 东北大学软件学院，辽宁 沈阳 110000）

**摘　要：**保留格式加密（format-preserving encryption，FPE）具有加密后数据格式和数据长度不变的特点，不会破坏数据格式约束，非常适用于数据脱敏领域。当前保留格式加密方案存在多次调用Cycle-Walking结构的不确定性问题或依赖模运算，导致原始消息空间与对称密文消息空间不对等的问题。基于动态有限域提出了基于动态有限域上的保留格式加密（dynamic galois field-format-preserving encryption，DGF-FPE）算法，对多字段的格式构建可以覆盖原始消息空间的有限域集合，在动态有限域中进行数据映射得到符合格式的FPE密文。解决了传统方法依赖Cycle-Walking或模运算方法的问题，并进行了相关实验。实验结果表明，经过DGF-FPE算法加密后的数据集的信息熵接近理想情况下的信息熵，证明了该算法具备充分的安全性。

**关键词：**保留格式加密；数据脱敏；动态有限域；对称密码算法；信息熵
**中图分类号：**TN918
**文献标志码：**A
**doi:** 10.11959/j.issn.1000−0801.DXKX250466

# Format-preserving encryption algorithm based on dynamic galois field

Yuan Heqing[1], Shen Jinshang[1], Wang Qiang[2]
1. Shandong Xinchao Information Technology Co., Ltd., Jinan 250100, China
2. Software College, Northeastern University, Shenyang 110000, China

**Abstract:** Format-preserving encryption (FPE) is characterized by its ability to maintain data format and length post-encryption while adhering to format constraints, rendering it highly applicable for data anonymization purposes. Current FPE schemes are faced with uncertainties in multiple calls to Cycle-Walking structures or dependencies on mod operations, leading to issues of asymmetry between the original message space and the symmetric ciphertext message space. The dynamic galois field-format-preserving encryption (DGF-FPE) algorithm based on the dynamic galois fields was proposed, which constructed a set of galois fields for multi-field formats that could cover the original message space. Data mapping was performed within the dynamic galois fields to obtain FPE ciphertext that conformed to

收稿日期：2025−07−22；修回日期：2025−12−29
通信作者：王强，wangqiang1@mail.neu.edu.cn
基金项目：国家自然科学基金资助项目（No.62202090，No.62173101）；中央高校基本科研业务费资助项目（No.N24170069）
**Foundation Items:** The National Natural Science Foundation of China (No.62202090, No.62173101), Research Funds for Central Universities (No.N24170069)

the format. This approach was addressed to the issues associated with traditional methods that relied on Cycle-Walking or mod operations, and relevant experiments were conducted. Empirical results reveal that datasets encrypted via DGF-FPE exhibit information entropy levels approaching theoretical maxima, thereby substantiating the algorithm's security profile.

**Key words:** FPE, data anonymization, dynamic galois field, symmetric encryption algorithm, information entropy

## 0 引言

随着信息技术的进步和大数据应用的发展，个人隐私信息保护问题越来越受到人们的重视。传统的个人信息保护方式基于信任边界，对可信的人赋予访问权限，对试图绕过信任边界非法获取权限的恶意威胁行为进行检测和阻断，这种保护方式可以在一定程度上预防非授权人员的数据访问。但据统计，授权人员恶意或无意的泄露行为是敏感数据泄露的主要途径，如内部人员滥用权限、非授权人员盗用身份等[1]，有研究表明，国内的信息泄露事故中85%是内部人员引起的[2]。

过去常用的数据脱敏手段包括替换、截断、偏移、加密、掩码和取整等[3]。这些方法从不同维度对敏感数据实施去隐私化处理，在一定程度上保护了隐私信息，但是传统的数据脱敏手段，一方面仍然是以真实的数据为基础的，同样会泄露部分隐私信息；另一方面无法还原为真实数据，在应用场景上有较大的限制。

在此背景下，2009年，Bellare等[4]从密码学理论的角度对保留格式加密（format-preserving encryption，FPE）进行了深入的研究，提出了FPE的密码学定义和安全目标。FPE是一种特殊的对称密码，基础模块是对称密码和伪随机函数，因其可在加密明文的同时保证密文与明文格式一致，已成为密码学和隐私保护领域的一个新兴分支。FPE的核心目标是保证加密后的密文具有和明文相同的格式、长度和数据类型。例如，保留格式加密后的身份证号仍为18位数字。相对于数据脱敏领域的其他技术，FPE在数据库加密、支付卡行业安全等领域具有独特的优势[5]。

## 1 研究现状

### 1.1 研究意义

FPE的特性特别适用于数据脱敏领域，传统的数据脱敏方法（如掩码、替换、哈希等）可能破坏数据格式，存在数据无法直接用于测试和影响系统兼容性等问题。FPE技术通常基于对称密码算法实现，保证了数据机密性的同时还具有较快的加密速度，又解决了传统密码算法改变数据格式的问题，可以直接应用于数据库等系统中。FPE作为加密技术还可以保证密文数据可以被解密回原始数据，在需要动态解密的场景中非常适用。

FPE技术的核心优势在于通过密码学手段实现格式无损加密，加密后的数据严格保留原始长度、字符集、校验规则等特征，可以无缝替代真实数据直接用于测试环境、数据分析或跨系统交互。同时，FPE特别适用于对轻量级数据的加密，例如对数据包头部进行加密，因此也可以用于计算机网络层的隐私保护[6]。与传统脱敏方法不同，FPE基于对称密码算法，既保障了数据的机密性又支持动态解密的灵活性，显著地提升了数据处理的灵活性与安全性。

### 1.2 FPE技术背景

当前保留格式加密方案大多基于Prefix、Cycle-Walking和Generalized-Feistel方法进行扩展和改进[7]，存在多次调用Cycle-Walking结构的不确定性问题，同时处理问题域的方法有一定的局限性。

Prefix 方法基于映射的思路将原始消息空间的每个元素都映射到新的消息空间并重新排序，排序后进行加密。这种方法的问题在于，如果明文空间的数据量极大（$2^{20}$以上）时，完成对消息空间的排序需要的计算资源也会大幅增加。

Cycle-Walking 方法为解决密文不在原始消息空间内的问题提供了一种通用的解决思路。加密的原理是利用基础分组密码（高级加密标准（advanced encryption standard，AES）或三重数据加密标准（triple data encryption standard，3DES）等）对中间输出值不断地进行处理，直至结果在正确的格式当中[8]。但这种方式会导致算法需要调用多轮 Cycle-Walking，造成了加密输出的不确定性[9]。

Generalized-Feistel 方法的核心思想有两点：一是使用 Feistel 结构调用基础分组密码算法构造适用于消息空间大小的分组密码算法；二是使用 Cycle-Walking 方法或模运算将加密所得消息映射到指定的消息空间。

Generalized-Feistel 方法的第二点导致了加密后仍须进行额外的计算。使用模运算方法，实际上使原始消息空间小于对称密文的消息空间，那么在解密时就存在需要判定 FPE 密文到底是对应哪个对称密文的问题，这就需要设计算法之外的判定机制。而使用 Cycle-Walking 方法又会导致加密次数和使用资源的不确定性。

基于这 3 种方法，Bellare 等[4]提出了检索增强的编码器（retrieval-augmented text encoder，RtE）模型，RtE 模型是一种相对比较通用的模型，基本思想是先排序后加密，将复杂问题域上的 FPE 问题转化为元素与索引的对应关系以及整数 FPE 问题。2010 年，Bellare 等[10]又提出了 FFX 模型，通过将固定字符表与其索引表建立双射关系，将字符串中的每个字符编码为数字串，对数字串进行 Feistel 加密运算，从而实现对消息空间上的保留格式加密。RtE 模型和 FFX 模型的加密过程都可以分解为算法初始化、对称加解密、数据映射 3 个部分。

### 1.3　国内外研究现状

在国内，丁建立等[1]将 FPE 技术和泛化技术结合，提出了泛化 FPE 模型，使攻击者无法建立明密文间明确的映射关系，从而保护明文数据的安全。李敏[7]基于保留格式加密技术，针对具有分段约束特征的编号类数据提出了相应的算法。王浩等[9]提出了一种基于数值的 GFPE 算法，通过构造加法群和有限域，同时使用 S 盒进行变换处理，实现对明文的混淆和扩散，但未对算法的随机性进行验证。刘俊等[11]提出了混合格式加密算法 IFX，通过建立多个编码表进行 FPE，并对编码表大小对加密时间的影响进行了实验，结果显示 IFX 不适合消息空间较大的情况。杨庆等[12]基于截断表设计了一种保留格式加密的方案，在 Feistel 结构必需的明文 $P$、密钥 $K$ 和密文 $S$ 之外，随机生成一个符合格式的数据 $R$，计算一个额外存储值 $C=R\oplus S$，将 $R$ 作为数据库中替换原数据的 FPE 后的数据。

国外的成果大多基于工程实践或结合其他技术[13-17]。Majeed 等[14]基于 FPE 技术提出了一种文本隐写系统，嵌入过程之前使用 FPE 技术加密秘密信息。Bansal 和 Garg[15]考虑当前在身份认证系统中，基于生物特征的认证数据在数据库中明文存储存在较高的风险，采用滤波器和 FPE 技术对生物特征进行加密的系统设计。Bellare 和 Huang[16]提出的基于身份的格式保留加密（identity-based format-preserving encryption，IB-FPE）通过身份密钥派生技术，引入基于身份的可调分组密码，实现在终端设备被攻破的情况下，所引起的安全风险也被限制在该终端设备当中。Jang 和 Lee[17]结合 FPE 技术加密后不改变原文格式的特点，提出了一种对图像具有隐私信息的部分进行 FPE 方案，通过精准定位和 FPE 隐私区域，可以在保持原始数据格式与图像可用性的同时保证隐私信息

不被泄露。

另外，有一些研究将FPE技术应用在硬件设备当中，旨在进一步扩大FPE技术的应用场景。Perez等[18]提出了一种CTR-MOD结构，基于分组密码的CTR工作模式和模运算，在FPGA上实现高速的保留格式加密。Kim等[19]在微处理器AVR和ARM上面，基于掩码技术设计了数据脱敏方案，并将之视为一种FPE方案。Agbeyibor等[20]评估了传统的FPE算法在关键基础设施上的适用性，通过在FPGA芯片上实现FF1、FF2和FF3算法评估了FPE在芯片中的安全性和硬件开销，评估结论认为FPE算法适用于关键基础设施当中。

### 1.4 字段加密问题

在本文所面向的保留格式加密的问题中，需要解决非独立字段的问题。这一问题的影响在于以下两个方面。

1）将独立字段和受其影响的非独立字段视为整个大的独立字段时，会造成排序或映射过程的困难，将提高排序所使用的计算资源或映射表的空间复杂度。

2）将独立字段和非独立字段分开计算时，受独立字段的影响，非独立字段的消息空间大小会有所变化，如果不加以处理，也可能出现前文中提到的对称密文消息空间与原始消息空间不匹配的问题，进而需要使用Cycle-Walking[21]结构或模运算方法。

### 1.5 本文的工作

本文所提的基于动态有限域上的保留格式加密（dynamic galois field-format-preserving encryption，DGF-FPE）算法，建立在RtE模型的基础上，对各个独立字段进行排序然后加密。通过动态有限域机制保证加密后的数据仍在原有的消息空间当中，解决了Generalized-Feistel方案[22]加密后仍须进行模运算或Cycle-Walking方法才能进入原始消息空间的问题。

## 2 基于动态有限域的FPE模型

数字特征数据除了具有基本数据类型的性质，还具有特定的自定义数据的完整性和约束性，例如，身份证号码、社保号码及电话号码等。这类数据可以分成若干的子集，每个子集都具有某些特定的构造规则。$Z$为特征数据，可以分成$n$个数据段，$Z_i$为子集，如式（1）所示。

$$Z=\bigcup_{i=1}^{n} Z_i \tag{1}$$

本文基于分段加密的思想构建动态有限域模型，结合分组密码或序列密码，实现保留格式加密算法。

### 2.1 有限域定义

本文构建的有限域计算的具体规则以$\mathrm{GF}\left(2^8\right)$域为例进行展示，后续使用的有限域计算都依据本节所展示的规则。

1）该域“×”和“⊕”运算满足交换律和结合律，如式（2）所示。

$$\begin{aligned}&\{11\}\times\{1010\}=\{1010\}\times\{11\},\\&\{11\}\times\{1010\}=\{01\}\times\{1010\}\oplus\{10\}\times\{1010\}\end{aligned} \tag{2}$$

2）域内加法⊕均为模2加运算。

3）域内除了0以外的所有元素$a$都有对应的乘法逆元$b$，使$a\times b$的结果可被本原多项式整除，即$a\times b\equiv 1 \bmod \left(m(x)\right)$。

4）乘法运算都在$\mathrm{GF}\left(2^8\right)$域内运算，运算的结果不能超出$\mathrm{GF}\left(2^8\right)$域的上限。

关于有限域乘法运算，设{02}与域内元素$a$相乘展示其计算过程。将元素$a$视为一个多项式，{02}表达为$x^1$，计算过程如式（3）所示。

$$\begin{aligned}\{02\}\times a=&\left\{\left(a_7x^8+a_6x^7+a_5x^6+\right.\right.\\&\left.\left.a_4x^5+a_3x^4+a_2x^3+a_1x^2+a_0x^1\right)*\left(x^1\right)\right)\\&\bmod\left(x^8+x^4+x^3+x^1+1\right),\end{aligned} \tag{3}$$

$$\{02\}\times a=\left\{\left(a_7x^8+a_6x^7+a_5x^6+a_4x^5+a_3x^4+a_2x^3+a_1x^2+a_0x^1\right)\right\}\ \mathrm{mod}\left(x^8+x^4+x^3+x^1+1\right)$$

其中，mod 中的$x^8+x^4+x^3+x^1+1$是$\mathrm{GF}\left(2^8\right)$域的本原多项式。将式中超出上限的$\left\{a_7x^8\right\}$部分，根据本原多项式转化成$\left(a_7x^4+a_7x^3+a_7x^1+a_7\right)$得到式（4）。

$$\{02\}\times a=\left\{\left(a_7x^8+a_6x^7+a_5x^6+a_4x^5+\ +\left(a_3+a_7\right)x^4+\left(a_2+a_7\right)x^3+a_1x^2+\left(a_1+a_7\right)x^1+a_7\right)\right\}\ \mathrm{mod}\left(x^8+x^4+x^3+x^1+1\right)\tag{4}$$

其中，$a_6x^7+a_5x^6+a_4x^5+\left(a_3+a_7\right)x^4+\left(a_2+a_7\right)x^3+a_1x^2+\left(a_0+a_7\right)x^1+a_7$即为相乘得到的计算结果。

## 2.2　符号说明

本节数学运算所用的变量名称较多，符号说明见表1。

**表1　符号说明**

| 符号 | 说明 |
|---|---|
| $R_{\mathrm{ini}}$ | 字段的原集合名称 |
| range | 字段对应的消息空间的大小 |
| $Z_i$ | 字段 |
| $A_i$ | 字段的消息空间 |
| $E(\cdot)$ | 对称密码算法 |
| $\mathrm{GF}_{\mathrm{mul}}\left(a\times b,2^n\right)$ | 在有限域$2^n$域中，进行$a\times b$乘法 |
| $\mathrm{mod}\ (a,b)$ | 模运算，计算$a$除以$b$的余数 |
| $\mathrm{find}(\mathrm{domain},a)$ | 查找算法，查找domain域中$a$元素的序号 |
| $P$ | 原始数据 |
| $P_1$、$P_2$ | FPE加解密过程中的过程数据 |
| $C$ | 对称密码算法输出的密文数据 |
| $C_{\mathrm{FPE}}$ | FPE得到的符合格式的密文 |
| $K_{\mathrm{data}}$ | 对称密钥 |
| $R_k^i$ | 字段$Z^i$的（消息空间）的第$k$个有限域 |
| $R_1$ | 有限域选择所使用的随机数$R_1\in\left[0,2^{n-1}-1\right]$ |
| $R_2$ | 有限域选择所使用的随机数$R_2\in\left[0,\mathrm{range}-1\right]$ |

## 2.3　基于动态有限域的FPE模型框架

基于动态有限域的DGF-FPE算法解决了Generalized-Feistel方法加密后仍须进行模运算或Cycle-Walking方法才能进入原始消息空间的问题。Generalized-Feistel方法加密后仍须进行模运算以及需要额外的验证部分的问题，本质上是原始消息空间（即问题域）和对称加密后的消息空间不匹配所导致的。原始消息空间小于对称密文消息空间的结构，虽然在加密时没有问题，但在解密时一个FPE密文会映射到多个对称密文上，对称加密前后的消息空间变化的示意图如图1所示。

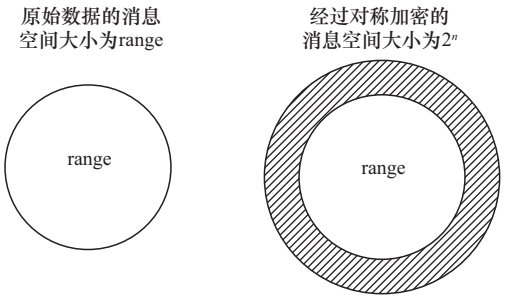


图1　对称加密前后的消息空间变化的示意图

图1右侧中，阴影区域代表着对称密文消息空间超出原始消息空间的数据集合，阴影区域的存在意味着每次FPE都有$\left(2^n-\mathrm{range}\right)/2^n$的概率加密结果不满足格式的要求。

这里需要指明动态有限域的概念，动态有限域概念的本质是动态的消息空间，借助随机数来动态生成大小规律的消息空间，解决对称加密前后的消息空间不匹配（无法一一对应）的问题。基于动态有限域的概念，本文设计的DGF-FPE算法，旨在通过随机数动态生成大小为2的整数幂的消息空间覆盖原始消息空间，进行保留格式加密。对称加密后的消息空间与动态有限域方法原理上的区别如图2所示。

如图2所示，在FPE过程中选择对称密文所在的集合，在此基础上构建动态有限域的映射规

则，将对称加密后的数据映射在动态的有限域中。因为有限域始终可以一一映射进原始消息空间中，所以不存在解密时一个FPE密文会映射到多个对称密文上的问题，且符合保留格式加密的要求，在计算量方面也明显小于3种传统的加密方法。

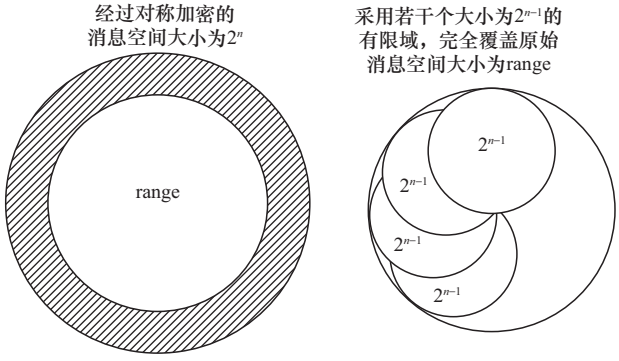


图2　对称加密后的消息空间与动态有限域方法原理上的区别

由于FPE技术作为加密技术的一种，本身就需要消除字段之间数据的关联性，所以，对于本文而言，为避免对称加密后的消息空间与原始消息空间可能不匹配的问题，将独立字段和受其影响的非独立字段视为整个大的独立字段，然后对各个独立字段独立进行FPE。基于动态有限域的保留格式加密模型如图3所示。

图3中，将保留格式的加密模型划分成3个部分：算法初始化部分、对称加解密部分、数据映射部分。其中，$x$表示初始明文，$x'$表示预处理过的明文，$S$表示初始密文，$y$表示处理后的密文。实际上，当前的FPE的各种算法都可以采用这一划分方式进行划分。

基于动态有限域的FPE流程如图4所示。

从图4中可以看出，在加密时首先对字段内数据统一进行排序和映射，得到原始消息空间。然后采用对称密码对明文数据$P$进行加密，得到对称密文数据。最后根据密文数据和随机数$R$，计算得到对称密文在动态有限域中对应的序号以及问题域中的密文$C_{\mathrm{FPE}}$。

### 2.4　明文加密具体流程

本文设计的DGF-FPE方法的加密过程可以简述为3个部分，如图3所示，以及具体的7个步骤，如图4所示。下面将论述3个部分和7个步骤的对应关系。

第1部分算法初始化，此部分对字段进行排序后建立映射表，对应下述过程的步骤1~2。需要指明的是，对于整个数据集而言，排序工作只需要执行一次。

第2部分对称加解密，使用对称密码算法对映射后的数据进行加密，并对加密后的结果使用第2部分选择的有限域和第1部分的逆映射表转换成符合格式的数据形式，对应下述过程的步骤3。

第3部分数据映射，对各个字段映射后的数

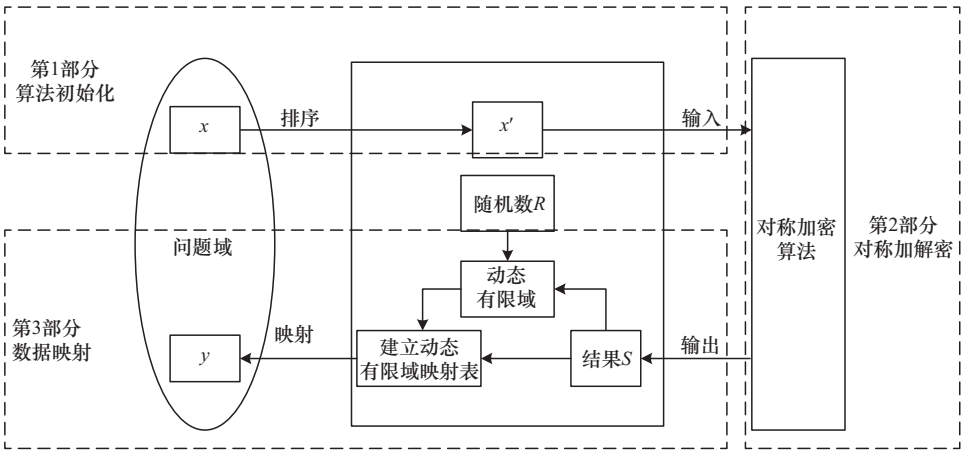


图3　基于动态有限域的保留格式加密模型

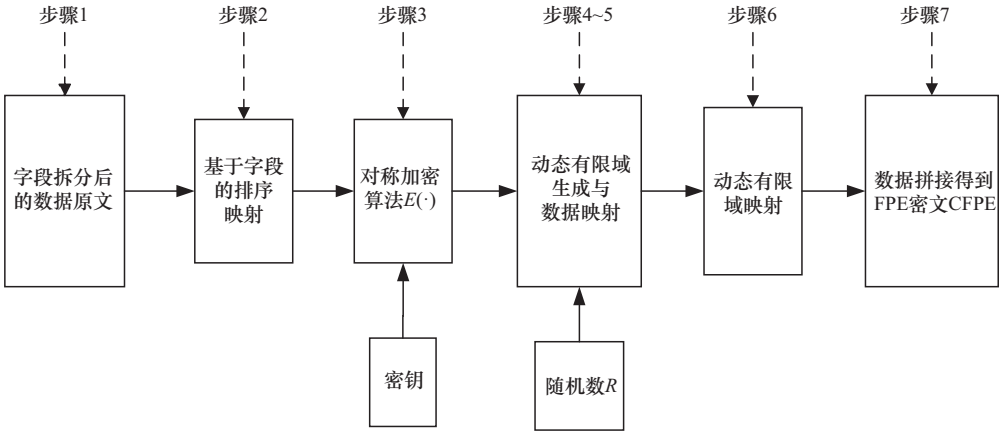


图4　基于动态有限域的FPE流程

据，根据随机数$R_1$动态地生成有限域，有限域集合中得到一个包括对称加密的密文数据$C$的有限域，并根据此对密文数据$C$进行映射，得到字段原始消息空间内的元素作为FPE的结果，对应下述过程的步骤4~7。

在图4中展示了本文设计的DGF-FPE算法流程，7个步骤的具体内容如下。

**步骤1**　划分字段集合

根据格式划分出数量最多的集合，先对独立字段进行后续的操作。如身份证号：110999-20000101-668-8，会被划分为3个相互独立的集合和1位校验码。对独立集合：行政区划号、出生日期、出生顺序进行FPE操作；而校验码在其他17位数据完成FPE后根据密文数据生成对应的校验码。

将各个字段命名为$Z_i$，对应的消息空间命名为$A_i$。

**步骤2**　子集排序

设此字段的原集合名为$R_{\text{ini}}$，原集合大小为range，与变量$n$存在关系$2^{n-1}<\text{range}<2^n$。使用排序算法将每个子集的元素进行从0到$\text{range}-1$的排序后，映射到一个大小为$[0,\ \text{range}-1]$，名为$A_1$的集合中，元素名为$\{a_0,a_1,\cdots,a_{\text{range}-1}\}$。同时得到明文$P$在$A_1$集合中的序号，将之命名为$P_1$。

**步骤3**　对称加密

构建分组密码或流密码算法对明文$P_1$进行加密，并输出对称密文$C$。此时$C$的消息空间大小为$2^n$，将此消息空间命名为$R_C$。

**步骤4**　动态有限域生成

使用动态有限域的方法完成对称密文$C$到原始消息空间的映射，需要生成$2^n$个子集合，作为有限域用于生成后面的映射表。存在关系$R_0\cup R_1\cup\cdots\cup R_{2^n-2}\cup R_{2^n-1}=R_C$，相当于一组有限域空间的集合（动态有限域）用来完全覆盖对称密文消息空间$R_C$，此步骤的概念可以参照前文中对动态有限域定义的描述和图2对动态有限域的展示。

$k$的值为$k\in[0,\text{range}-1]$，生成的$2^n$个子集合，各个集合的组成元素如式（5）所示。

$$R_k=\{x\in Z|[(k\leqslant x\leqslant 2^{n-1}+k)\cup(x\leqslant k+2^{n-1}-(\text{range}-1))\cap[0,2^n-1]\tag{5}$$

1个$R_k$相当于1个集合，存在关系$R_0\cup R_1\cup\cdots\cup R_{\text{range}-2}\cup R_{\text{range}-1}=A_1$，相当于一组有限域集合可以完全覆盖集合$A_1$。

**步骤5**　数据映射

在步骤4生成的$2^n$个子集合中，包含对称密文$C$的集合只占其中的1/2，将这些集合根据步骤4中$R_k$的序号，重新从0到$2^{n-1}$进行排序。找到最

后一位元素为对称密文$C$的有限域，将其命名为$D$域，其组成元素如式（6）所示。

$$\begin{cases} D=\left[C-2^{n-1}+1:\ C\right],\ C-2^{n-1}\geqslant 1 \\ D=\left\{\left[2^{n-1}+C+1:\ 2^{n}\right],\left[1:\ C\right]\right\},\ C-2^{n-1}<1 \end{cases} \tag{6}$$

将随机数$R_1$映射进$D$域中，得到对称密文$C$与随机数$R_1$在集合中相隔的元素数量为$P_2$，并生成tag标识。$P_2$和tag的计算如式（7）和式（8）所示。

$$\begin{cases} P_2=2^{n-1}-\text{find}\left(D,2^{n-1}+R_1\right),\ C>2^{n-1}+R_1, \\ P_2=2^{n-1}-\text{find}\left(D,1+R_1\right),\ C\leqslant 2^{n-1}+R_1 \end{cases} \tag{7}$$

$$\begin{cases} \text{tag}=1,\ C>2^{n-1}+R_1, \\ \text{tag}=0,\ C\leqslant 2^{n-1}+R_1 \end{cases} \tag{8}$$

由于$D$域的最后一位元素必然是对称密文$C$，所以在$R_1$已知的情况下，可以根据随机数$R_1$在$D$域的序号$P_2$，逆推出$D$域的范围和对称密文$C$的值。

从外部输入一个随机数$R_1$，对$P_2$进行有限域的乘法计算，计算得到的index作为对称密文$C$在$D$域进行保留格式映射所使用的密文编号，计算过程如式（9）所示。

$$\text{index}=\text{GF}_{\text{mul}}\left(P_2\times R_1,2^{n-1}\right) \tag{9}$$

**步骤6** 动态有限域映射

同时，随机数$R_2$也用于生成动态有限域。$R_2$和index直接相加得到$\text{FPE}_{\text{index}}$，$\text{FPE}_{\text{index}}$的分布空间为$\left[0,\text{range}-1\right]$，计算过程如式（10）所示。

$$\text{FPE}_{\text{index}}=\text{mod}\left(R_2+\text{index},\text{range}\right)+1 \tag{10}$$

根据步骤2生成的集合$R_{\text{ini}}$，将序号值$\text{FPE}_{\text{index}}$映射进去，将序号对应的元素作为FPE的字段输出结果。

**步骤7** 数据拼接

对明文$P$各字段进行步骤2~6的操作，将各个字段加密后的数据进行拼接，得到符合FPE要求的密文$C_{\text{FEP}}$。

### 2.5 密文解密具体流程

本文所提方法DGF-FPE的解密过程同样可以简述为3个部分。但解密的顺序与加密的3个部分不同。

第1部分算法初始化，对字段进行排序后建立映射表，对应下述过程的步骤1~2。此部分与加密的第1部分完全相同。

第2部分数据映射，对各个字段映射后的数据生成若干个有限域，根据随机数$R_1$、$R_2$从有限域集合中确定对称加密的密文数据$C$所在的有限域，并对密文数据$C_{\text{FEP}}$进行映射，得到对称加密的密文数据$C$，对应下述过程的步骤3。此部分对应加密的第3部分。

第3部分对称加解密，使用对称密码算法对密文数据$C$进行解密，并对加密后的结果使用第1部分的逆映射表转换成符合格式的数据形式，最后根据第1部分的映射表解密出原始数据。对应下述过程的步骤4~6。

**步骤1** 划分字段集合

此部分与加密部分的步骤1完全一致，不再赘述。

**步骤2** 子集排序

此部分与加密部分的步骤2完全一致，不再赘述。

**步骤3** 动态有限域逆映射

根据$\text{FPE}_{\text{cipher}}$和步骤2的排序结果，得到$\text{FPE}_{\text{index}}$。根据随机数$R_2$和$\text{FPE}_{\text{index}}$计算出index。最后根据随机数$R_2$和index计算出对称密文$C$与随机数$R_1$的乘法逆元$R_{1\,\text{inv}}$在集合中相隔的元素数$P_2$。计算过程如式（11）和式（12）所示。

$$\text{index}=\text{mod}\left(\text{FPE}_{\text{index}}-R_2,\text{range}\right)-1 \tag{11}$$

$$P_2=\text{GF}_{\text{mul}}\left(\text{index}\times R_{1\,\text{inv}},2^{n-1}\right) \tag{12}$$

由于$D$是顺序排列，且密文数据$C$永远对应$D$的最后一位序号，所以，根据$P_2$和随机数$R_1$可以计算出密文数据$C$的值，计算过程如式（13）

所示。

$$C=\mathrm{mod}\left(\mathrm{tag}\times\left(2^{n-1}\right)+P_2+R_1+1,2^n\right)\quad(13)$$

**步骤4**　对称解密

构建分组密码或流密码算法对密文数据$C$进行解密，并输出对称解密的结果数据$P_1$。

**步骤5**　数据映射

根据明文$\mathrm{data}_{\mathrm{index}}$所代表的序号，查找其在$R_{\mathrm{ini}}$中所代表的元素，此元素即对应字段中的原始明文数据。

**步骤6**　数据拼接

对各字段的集合进行步骤2～6的操作，将各个字段解密后的数据进行拼接，得到对称密码算法加密前的原始数据$P$。

## 3　实验与结果分析

本文使用的实验环境如下：MATLAB R2016a、CPU为Intel（R）Core（TM）i5-10300H，内存16 GB，操作系统为Windows 10。基于动态有限域的保留格式加密算法进行正确性、安全性和效率分析实验，以身份证号码为例。所使用的对称密码为祖冲之（ZUC）算法，设定初始密钥与初始向量分别为16’h3d4c4be96a82fdaeb58f641db17b455b和16’h84319aa8de6915ca1f6bda6bfbd8c766。ZUC算法作为流密码每次固定生成一个32位的密钥字，如果该字段的消息空间小于$2^{32}$，则生成一个密钥字用于加密，如果消息空间超过$2^{32}$，则每增加$2^{32}$就增加一个密钥字用于加密。

本节将基于身份证号码进行保留格式加密和解密过程，对本文所提的DGF-FPE算法的正确性和统计特性进行验证。本文将身份证号视为4个部分，分别是行政区划号、出生日期、出生日期顺序号、校验码，示例如下：

某人的身份证号110105********002X被拆分为110105-********-002-X。本节将对其中前3个部分是独立的消息字段进行FPE加密实验，最后一位校验码由前面的17位数据计算生成。

### 3.1　算法正确性验证

实验随机生成30 000条身份证号数据作为实验数据，加密处理部分实验数据，FPE加解密数据示例见表2。

### 3.2　安全性分析

保留格式加密是一种特殊的对称密码，基础模块是分组密码和伪随机函数，因此它的一个重要的安全目标是伪随机性。2002年，Rogaway等[23]首次描述FPE的安全性，认为标准FPE安全就是伪随机置换（pseudo random permutation，PRP）安全。也就是说，FPE的本质是在特定消息空间内的伪随机置换。当前的FPE基本都基于对称加密算法，可以保证加密过程中的对称密文具有可靠的随机性。

表2　FPE加解密数据示例

| 原文 | $R_1$ | $R_2$ | Key_word | FPE密文数据 | FPE解密数据 |
|---|---|---|---|---|---|
| 411524********0015 | 511 | 103 | E3C4D158<br>14F1C272<br>3279C419 | 510823********6148 | 411524********0015 |
| 411602********61 | 511 | 103 | E3C4D158<br>14F1C272<br>3279C419 | 220105********1355 | 411602********61 |
| 320581********996X | 511 | 103 | E3C4D158<br>14F1C272<br>3279C419 | 230522********1724 | 320581********996X |
| 320613********0018 | 511 | 103 | E3C4D158<br>14F1C272<br>3279C419 | 511825********6145 | 320613********0018 |

3.2.1 PRP安全

保留格式加密以PRP安全为安全目标，其安全性等效于在原始消息空间进行随机置换。本文所提的DGF-FPE算法，使用ZUC算法作为FPE的伪随机函数，对数据的各个独立字段进行ZUC加密后，使用随机数进行动态有限域的随机映射。

Bellare和Cash[24]基于游戏模型，进一步定义了适用于FPE一般定义的PRP安全目标。通过攻击者与挑战者之间的博弈，对攻击者区分保留格式加密和随机置换的能力进行量化。首先，随机选择密钥$k$、待猜测的布尔值$b$和随机置换$P$。$\mathrm{PRP}_{\varepsilon\mathrm{FPE}}$拥有加密预言机Oracle（Enc），该加密预言机在收到攻击者A的加密查询后，会随机选择保留格式加密或随机置换对攻击者A输入的明文数据进行处理，并返回处理过的密文，具体工作过程如下：如果$b=1$，加密预言机Oracle（Enc）使用本文设计的DGF-FPE算法对攻击者输入的明文数据进行处理并输出处理后的密文，即进行εFPE响应（εFPE表示使用本文提出的DGF-FPE算法处理数据的过程）；如果$b=0$，加密预言机Oracle（Enc）使用随机置换对攻击者输入的明文数据进行处理并输出处理后的密文，即进行$P$响应（$P$表示使用随机置换处理数据的过程）。攻击者A的目标是：在查询加密预言机Oracle（Enc）一定次数后，给出一个判定，Oracle（Enc）使用的是εFPE还是$P$来响应攻击者A的请求，即输出一个判定值$b'$。如果$b'=b$，说明攻击者A判定成功，相反则失败。A在该游戏的优势如式（14）所示。

$$\mathrm{Adv}^{\mathrm{PRP}}_{\varepsilon_{\mathrm{DGF-FPE}}}(\mathrm{A})=2\Pr\left(\mathrm{PRP}^{\mathrm{A}}_{\varepsilon_{\mathrm{DGF-FPE}}}\Rightarrow\mathrm{true}\right)-1 \quad (14)$$

当攻击者在该游戏中具有的优势不明显时，则该FPE方案εFPE达到了PRP安全。

接下来，本文将基于FPE对身份证号的出生日期顺序号部分进行3次实验，分析不同的随机数$R_1$和$R_2$生成规则下，DGF-FPE算法的统计分布情况，并论证本文所提的基于动态有限域的保留格式加密的PRP安全。

3.2.2 频率直方图

频率直方图是对小范围的数据集、各明文的样本值和密文的样本值出现的频率进行统计和分析得到的图像，通过频率直方图可以看出整个数据集中各密文的样本值的分布情况，也可推断出加密的效果，常用于图像加密领域。由于FPE中，独立消息字段的消息空间较小（一般不超过$2^{32}$），适用于频率直方图展示FPE对单个独立消息字段的加密效果。

3.2.3 信息熵

信息熵是信息理论中的一个重要概念，用于描述信息或随机变量的不确定性。信息熵越大，不确定性就越大。对于本文的研究而言，信息表示原始消息空间中每条信息出现的概率。信息熵的计算式如式（15）所示。

$$H(x_i)=-\sum_{i=1}^{N}p(x_i)\mathrm{lb}\,(x_i) \quad (15)$$

身份证号的出生日期顺序号，其样本值分布在[1, 999]区间中。理想情况下的FPE所有样本出现的概率都为1/999，信息熵为lb(999) = 9.964 3。信息熵越接近9.964 3，则说明随机性越好，越接近PRP安全。

3.2.4 出生日期顺序号FPE加密实验

实验随机生成30 000条身份证号数据作为实验数据，对其进行加密处理实验，并采用频率直方图和信息熵作为安全指标。

**实验1** 随机数$R_1$和$R_2$采用固定常数的情况下，随机数取固定常量时的频率直方图如图5所示，展示了原始消息空间中$\mathrm{FPE}_{\mathrm{cipher}}$的分布情况。

图5中可以看出，当$R_1$和$R_2$采用固定数值的情况下，$\mathrm{FPE}_{\mathrm{cipher}}$仅在$[0, 2^{n-1}]$的域内实现了充分的随机分布，可以明显看出并非在原始消息空间

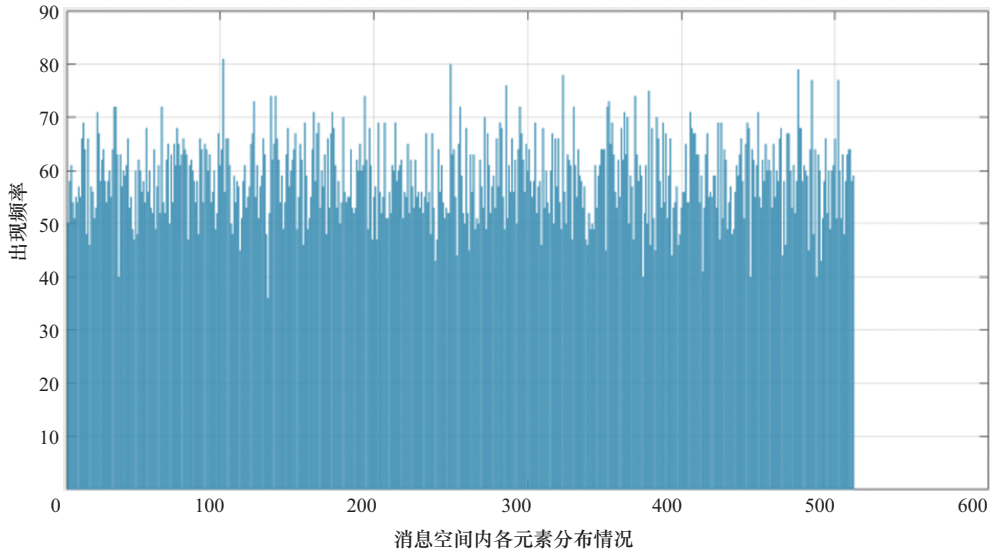


图5　随机数取固定常量时的频率直方图

$R_{ini}$ 中随机分布，且信息熵为 8.988 3 与理想值 9.964 3 仍有较大差距，无法满足 PRP 安全。

**实验 2**　采用模运算生成随机数 $R_1$ 和 $R_2$ 的情况中，采用模运算生成随机数时的频率直方图如图 6 所示，展示了原始消息空间中 $FPE_{cipher}$ 的分布情况。

图 6 中可以看出，当 $R_1$ 和 $R_2$ 采用模运算生成的情况下，$FPE_{cipher}$ 的分布情况明显不是均匀分布的，信息熵为 9.689 3 与理想值 9.964 3 仍有一定的差距，无法满足 PRP 安全。

**实验 3**　根据表 1 的定义，在定义范围内均匀生成随机数 $R_1$ 和 $R_2$ 的情况下，均匀生成随机数时的频率直方图如图 7 所示，展示了原始消息空间中 $FPE_{cipher}$ 的分布情况。

图 7 中可以看出，当 $R_1$ 和 $R_2$ 采用均匀生成随机数的情况下，$FPE_{cipher}$ 的分布情况是相对均匀的，且信息熵为 9.936 4 非常接近理想值 9.964 3，可以满足 PRP 安全。

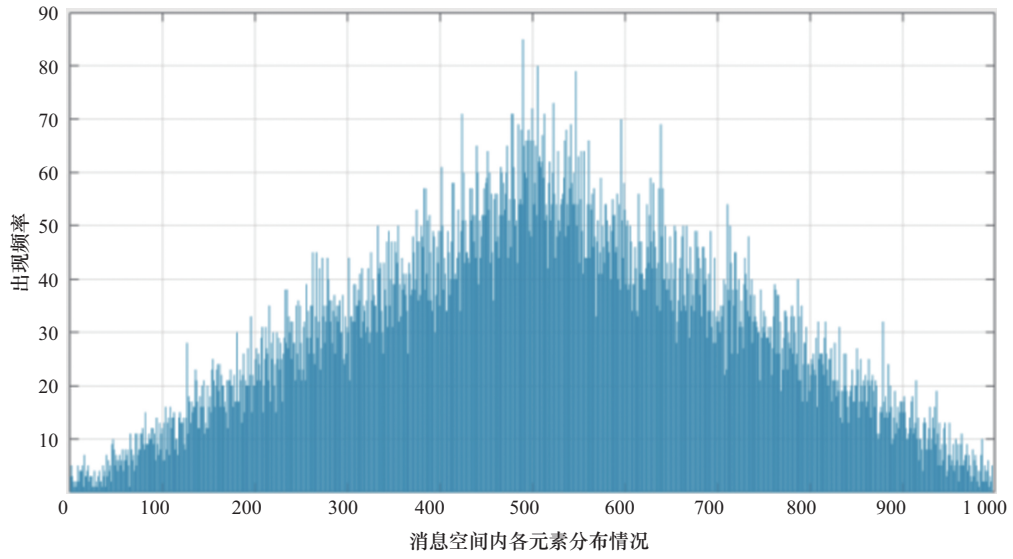


图6　采用模运算生成随机数时的频率直方图

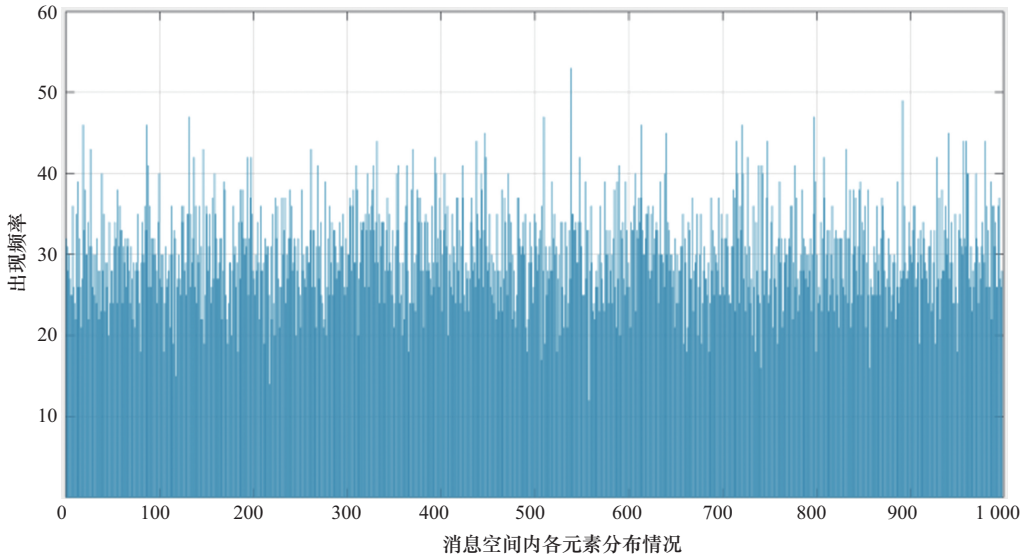


图7　均匀生成随机数时的频率直方图

实验1~3的对比表明，采用随机分布生成随机数的方法可以使FPE输出的密文均匀地随机分布在原始消息空间当中且信息熵接近理想值，难以区分FPE与随机置换的区别。可以得出结论，本文基于DGF-FPE算法结合ZUC算法实现的FPE算法可以满足PRP安全。

### 3.3　效率分析与特征对比

本文所实现的DGF-FPE算法是基于RtE模型实现的，除了RtE模型所必需的排序算法和查表计算外，仅使用一次有限域运算和一次可逆的模运算即可完成，相对于对称加解密所需的计算资源明显较少，具有较高的加密效率。

本文提出的DGF-FPE算法，可以基于流密码或分组密码作为基础对称密码算法构建。本文从算法特征的一些指标进行对比，对比特征包括基础对称密码算法、适用数据类型、密钥长度、关键步骤、数据映射阶段对Cycle-Walking或模运算方法的依赖程度、效率几方面，保留格式加密算法实现方案对比见表3。

从表3中可以看出，DGF-FPE算法不依赖Cycle-Walking或模运算方法，在效率方面具备一定的优势，同时兼容分组密码算法和序列密码算法，灵活性更好。

在前文中提到过，实际上FPE模型可以分解为3个组成部分：算法初始化部分、对称加解密部分、数据映射部分。本文对复杂度的分析也将按此划分方式进行，下面先对3个组成部分进行分析。

表3　保留格式加密算法实现方案对比

| 指标 | FFSEM | FF3 | GFPE[9] | TT-FPE[12] | DGF-FPE |
|---|---|---|---|---|---|
| 基础对称密码算法 | AES | AES、3DES | AES | AES、3DES | ZUC、高级加密标准-计数器模式（advanced encryption standard-counter，AES-CTR） |
| 数据类型 | int | int | int、char | int、char | int、char |
| 密钥长度 | 128 bit | 128/64 bit | 128 bit | 128/64 bit | 128 bit |
| 关键步骤 | 轮运算 | 数据映射、轮运算 | 轮运算 | 轮运算 | 密钥流生成、数据映射 |
| Cycle-Walking或模预算 | 依赖高 | 依赖低 | 依赖高 | 不依赖 | 不依赖 |
| 效率 | 低 | 中 | 中 | 中 | 高 |

（1）算法初始化部分：此部分是将原始数据和原始消息空间映射进整数域的基础，需要建立一次映射表。

（2）对称加解密部分：基于对称密码算法实现对数据的加密，需要大量的数学变换，是算法中主要的计算部分，也是算法安全性的基础。值得一提的是，只要能完成对消息空间内数据的加密，对FPE而言选择哪个对称密码算法实现对称加解密部分是无关紧要的。

（3）数据映射部分：确保对称加密后的数据经过对称加密后可以被映射进原始消息空间当中，是使数据满足保留格式要求的必要组成部分。

在这3部分中，对称加解密部分需要大量的数学变换，因此占据了主要的计算资源和计算时间；其次是数据映射部分，数据映射实际上占据的资源并不多，但如果使用Cycle-Walking方法将有可能重复调用对称加解密部分，因此，数据映射部分也对计算复杂度和时间复杂度有较大的影响；算法初始化部分通常只需要建立一个映射表即可，根据字段拆分后映射表相对较小，占用资源也较小。因此，本文将对比使用的基于ZUC算法的DGF-FPE算法和AES、SM4算法和ZUC算法的用时情况，进而论证FPE算法的复杂度与其所使用的对称密码算法之间的关系。

用时情况比较如图8所示，绘制了在数据量逐渐增加时，基于ZUC算法的DGF-FPE算法和AES、SM4算法和ZUC算法的用时情况。从图8中可以看出，在数据量逐渐增加时，3种算法的用时情况呈线性增长，其中，ZUC算法和基于ZUC算法的DGF-FPE算法计算速度相对较快，也可以看出基于ZUC算法实现的DGF-FPE算法主要的计算量是由其3个组成部分中的对称加解密部分（ZUC算法）引起的。因此，推荐使用计算量较小的对称密码算法作为FPE的基础。

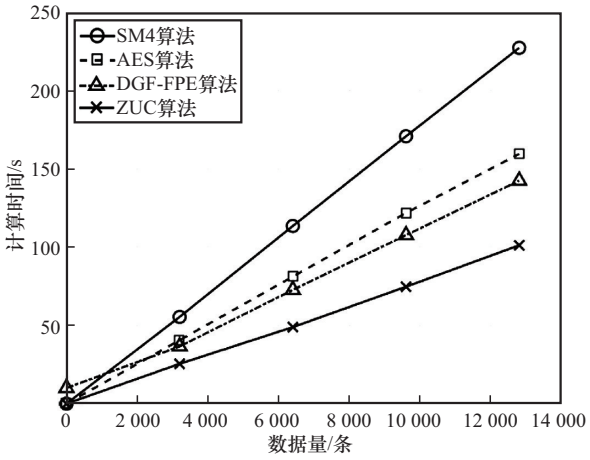


图8　用时情况比较

## 4　结束语

本文针对当前保留格式加密中Cycle-Walking和Generalized-Feistel方法无法保证密文正确落入格式消息空间的加密次数的问题，提出了基于动态有限域的保留格式加密模型，并将FPE技术拆分为算法初始化、对称加解密和数据映射3个部分。该方法采用动态有限域的方法，在原始消息空间和对称密文空间中建立若干个动态的有限域，通过对动态有限域进行操作，确保一次对称加密就能使密文满足格式要求，有效地提高了算法效率。

对所提的DGF-FPE算法设计了实验。通过对其频率直方图和信息熵进行分析，证明了该算法的安全性。在效率分析中指出了FPE算法的复杂度主要取决于所使用的对称密码算法，进行相关实验验证了DGF-FPE算法的计算效率并进一步论证了对称密码算法对FPE算法复杂度的影响，可以确定该算法可用于对大型数据集进行保留格式加密。

## 参考文献：

[1]　丁建立, 陈盼, 马勇. 基于泛化FPE加密的民航旅客信息动态脱敏方法研究[J]. 信息网络安全, 2021, 21(2): 45-52.
Ding J L, Chen P, Ma Y. Research on dynamic desensitization

method of civil aviation passenger information based on generalized FPE encryption[J]. Netinfo Security, 2021, 21(2): 45-52.

[2] You F, Zhang C, Cao Y, et al. Data masking system based on ink technology[C]//Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE). Piscataway: IEEE Press, 2018: 176-180.

[3] 仇壮丽, 王昱雯. 基于隐私保护的档案数据脱敏路径分析[J]. 山西档案, 2025, (4): 23-31.
Chou Z L, Wang Y W. Analysis of desensitization path for archive data based on privacy protection[J]. Shanxi Archive, 2025, (4): 23-31.

[4] Bellare M, Ristenppart T, Rogaway P, et al. Format-preserving encryption[C]//Proceedings of the Selected Areas in Cryptography: 16th Annual International Workshop, SAC 2009, Calgary, Alberta, Canada, August 13-14, 2009, Revised Selected Papers 16. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 295-312.

[5] Bhatt Z, Gupta V. Advance security technique for format preserving encryption[C]//Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2016, 1: 1-4.

[6] Micovic M, Radenkovic U, Vuletic P. Network layer privacy protection using format-preserving encryption[J]. Electronics, 2023, 12(23): 4800.

[7] 李敏. 保留格式加密技术应用研究[D]. 天津: 南开大学, 2012.
Li M.Research on the applications of format preserving encryption[D]. Tianjin: Nankai University, 2012.

[8] 刘哲理, 贾春福, 李经纬. 保留格式加密技术研究[J]. 软件学报, 2012, 23(1): 152-170.
Liu Z L, Jia C F, Li J W. Research on the format-preserving encryption techniques[J]. Journal of Software, 2012, 23(1): 152-170.

[9] 王浩, 张永平, 李同寒, 等. 一种数值型的保留格式加密算法[J]. 信息安全研究, 2023, 9(8): 745-753.
Wang H, Zhang Y P, Li T H , et al. A format-preserving encryption algorithm for numeric data[J]. Journal of Information Security Research, 2023, 9(8): 745-753.

[10] Bellare M, Rogaway P, Spies T. The FFX mode of operation for format-preserving encryption[J]. NIST Submission, 2010, 20(19): 1-18.

[11] 刘俊, 李泽昊, 苏国宇, 等. 保留格式加密技术在民航信息系统数据处理中的应用研究[J]. 计算机科学, 2019, 46(S1): 571-576.
Liu J, Li Z H, Su G Y, et al. Application of reserved format encryption technology in information processing of civil aviation information system[J]. Computer Science, 2019, 46(S1): 571-576.

[12] 杨庆, 田有亮, 熊金波. 基于截断表的保留格式加密算法TT-FPE[J]. 福建师范大学学报(自然科学版), 2025, 41(1): 68-76.
Yang Q, Tian Y L, Xiong J B. A truncationed table-based format-preserving encryption algorithm TT-FPE[J]. Journal of Fujian Normal University (Natural Science Edition, 2025, 41(1): 68-76.

[13] Habibi M J, Amrhar A, Gagne J M, et al. Security establishment in ADS-B by format-preserving encryption and blockchain schemes[J]. Applied Sciences, 2023, 13(5): 3105.

[14] Majeed M A, Sulaiman R, Shukur Z. New text steganography technique based on part-of-speech tagging and format-preserving encryption[J]. KSII Transactions on Internet and Information Systems (TIIS), 2024, 18(1): 170-191.

[15] Bansal V, Garg S. A cancelable biometric identification scheme based on bloom filter and format-preserving encryption[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(8): 5810-5821.

[16] Bellare M, Hoang V T. Identity-based format-preserving encryption[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2017: 1515-1532.

[17] Jang W, Lee S Y. Partial image encryption using format-preserving encryption in image processing systems for Internet of things environment[J]. International Journal of Distributed Sensor Networks, 2020, 16(3): 1550147720914779.

[18] Perez R A, Garcia B M, Sanchez A C, et al. A new method for format preserving encryption in high-data rate communications[J]. IEEE Access, 2020, 8: 21003-21016.

[19] Kim H, Sim M, Jang K, et al. Masked implementation of format preserving encryption on low-end AVR microcontrollers and high-end ARM processors[J]. Mathematics, 2021, 9(11): 1294.

[20] Agbeyibor R, Butts J, Grimaila M, et al. Evaluation of format-preserving encryption algorithms for critical infrastructure protection[C]//Proceedings of the International Conference on Critical Infrastructure Protection. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 245-261.

[21] Hobbelen D G E, Wisse M. Limit cycle walking[J]. InTech, 2007. DOI: 10.5772/4808.

[22] Nyberg K. Generalized Feistel networks[C]//Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security. Springer-Verlag, 1996. DOI: 10.1007/BFb0034838.

[23] Rogaway P, Bellare M, Black J, et al. OCB: a block-cipher

mode of operation for efficient authenticated encryption[C]// Proceedings of the Conference on CCS. New York: ACM Press, 2001. DOI: 10.1145/501983.502011.

[24] Bellare M, Cash D. Pseudorandom functions and permutations provably secure against related-key attacks[J]. Springer, Berlin, Heidelberg, 2010. DOI: 10.1007/978-3-642-14623-7_36.

[作者简介]

**元河清**（1988– ），男，山东新潮信息技术有限公司高级工程师，主要研究方向为密码学应用与网络安全。

**申锦尚**（1999– ），男，山东新潮信息技术有限公司工程师，主要研究方向为密码学应用、高速数据通信。

**王强**（1991– ），男，博士，东北大学软件学院副教授、硕士生导师，主要研究方向为云计算安全、安全多方计算、区块链。