基于中国剩余定理的强鲁棒性多级数据库图像水印嵌入方案

耿慧拯，郭斯栩，刘颖卿，粟栗，何申
（中国移动通信有限公司研究院，北京 100053）

摘　要：随着数据逐渐成为企业的一种资产形式，为了更好地发挥数据的价值，避免出现数据“孤岛”问题，结构化数据在企业内部和外部的共享场景日益增多。然而，由于现有的数字水印方案大多只能添加一次水印信息，如果重复添加，新水印会覆盖原有水印信息，导致无法实现对数据的多级溯源。针对上述问题，提出了一种基于中国剩余定理的多级数据库图像水印嵌入方案（multi-level database image watermark embedding scheme based on the Chinese remainder theorem，MIWC）。该方案以图像水印作为载体，通过Harr小波变换、布隆过滤器等方法对图像进行预处理，形成像素点布隆过滤器，并根据中国剩余定理的性质对水印信息进行秘密分片，实现对水印信息的分级逐条添加，进而对数据整个流转链路进行记录和溯源。功能性分析表明，该方案具备现有方案所缺乏的多级水印嵌入能力，具有较高的实际应用价值。实验结果显示，该方案具备可逆性，水印的嵌入与提取均表现出较高效率。同时，方案还具有较强的鲁棒性，能够抵抗针对水印的攻击。

关键词：数据库水印；数据溯源；中国剩余定理；Harr小波变换
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025181

A strong robust multi level database image watermark embedding scheme based on the Chinese remainder theorem

GENG Huizheng, GUO Sixu, LIU Yingqing, SU Li, HE Shen
China Mobile Communications Corporation Research Institute, Beijing 100053, China

Abstract: With data gradually becoming a form of asset for enterprises, structured data is increasingly being shared both internally and externally to better leverage its value and avoid data silos. However, due to the fact that most existing digital watermarking schemes can only add watermark information once, repeated additions will overwrite the original watermark information, making it impossible to achieve multi-level traceability of data. In response to the above issues, a multi-level database image watermark embedding scheme based on the Chinese remainder theorem (MIWC) was proposed. This scheme used image watermarking as a carrier, preprocesses the image through Harr wavelet transform, Bloom filter and other methods to form pixel Bloom filter, and secretly segmented the watermark

收稿日期：2024−07−19；修回日期：2025−07−30
通信作者：郭斯栩，13810202832@163.com

information according to the properties of the Chinese remainder theorem. The watermark information could be added step by step in a hierarchical manner, achieving the recording and traceability of the entire data flow chain. Functional analysis proves that the solution has the ability to embed multi-level watermarks that existing solutions do not have, and has high practical application value. Experimental results have shown that this scheme has reversibility and high efficiency in embedding and extracting watermarks. At the same time, the scheme has strong robustness, it can resist attacks against watermarks.

Key words: database watermark, data traceability, Chinese remainder theorem, Harr wavelet transform

0 引言

数字水印[1-2]技术的研究背景主要源于互联网和多媒体信息技术迅猛发展所带来的信息安全挑战。随着多媒体信息的广泛传播及其所具有的易于被修改的特性，版权归属、信息完整性和真实性等问题逐渐凸显。这些问题不仅损害了版权所有者的合法权益，还可能牵涉个人隐私、企业机密和国家安全等更为广泛的信息安全。

作为应对上述挑战的一种有效手段，数字水印技术在不引起原载体图像视觉失真且不影响其正常使用的前提下，将水印信息嵌入数字载体中。该技术不仅能用于版权保护，即通过在作品中嵌入代表版权所有者的水印信息，以确保在作品被篡改或盗用时能够判别版权归属，还能用于追踪盗版数字内容的来源，为打击侵权行为提供有力支持。此外，数字水印技术凭借其不可感知性、可验证性和不需要密码管制的特点，在众多数字版权保护标准中占据了举足轻重的地位。随着研究的不断深入，数字水印技术已从最初主要应用于静止数字图像，逐步拓展至视频、音频和文档等其他领域，为信息安全提供了更为全面、有效的保障。因此，对数字水印技术的研究不仅具有深远的现实意义，而且展现出广阔的应用前景。随着技术的持续进步和完善，数字水印将在维护信息安全、推动多媒体信息健康发展方面发挥更加关键的作用。

然而，数字水印在实际应用中面临着多重挑战，如易被攻击者识别并清除、对媒体质量与载体有特定要求，以及在不同媒体应用领域的适应性差异等。为了应对这些挑战，研究者们正努力提出各种改进和增强的数字水印方案。例如，通过引入对抗训练来强化水印的鲁棒性，利用多级水印、分级水印等技术提升水印的适应性和稳定性，以及结合数字水印与加密技术[3]为版权保护提供更为坚固的屏障。

为了防止恶意攻击者未经版权所有者许可就复制、篡改和传播受版权保护的信息，数据库所有者会在数据库中嵌入水印，以证明版权的归属或验证数据的完整性。可逆水印技术[4-5]作为数字水印领域的关键分支，一直以来都吸引着广大研究者的目光。在可逆数据库水印算法的研究中，文献[6]提出了一种基于直方图平移的经典算法，通过创建额外的冗余空间并遵循特定规则来嵌入水印信息。为了进一步完善上述文献的算法，文献[7]创新性地引入了遗传算法，通过算法训练来寻找最优的水印嵌入位置，结合分组密钥与直方图平移技术，开发出一种新颖的基于遗传算法的可逆水印方法。此外，为了提高水印的鲁棒性，众多新颖的数字水印方案被相继提出。例如，文献[8]提出了一种基于差分扩展的可逆数据库盲水印方案；文献[9]则利用遗传算法和差分扩展技术设计出可逆数据库水印；文献[10]在水印容量和数据失真方面进行了深入研究与改进，在可接受的失真范围内提升了水印的嵌入容量；文献[11]则将大容量、高容错的二维码理论应用于变换域图像水印算法，显著增强了水印的鲁棒性。

当前，随着数据逐渐成为企业的重要资产，为了最大化数据的价值并规避数据孤岛现象，结构化数据在企业内部及跨企业间的共享场景正日益增多。然而，现有的数字水印技术普遍面临一个限制，即它们大多仅支持单次水印信息的添加，一旦尝试重复添加，新水印往往会覆盖原有水印，导致无法实现数据的多级溯源。此外，在数据共享过程中，如何保护发布方的版权，并确保共享数据不被第三方非法泄露或倒卖，成为亟待解决的问题。鉴于上述挑战，本文提出了一种基于中国剩余定理的强鲁棒性多级数据库图像水印嵌入方案（multi-level database image watermark embedding scheme based on the Chinese remainder theorem，MIWC）。该方案巧妙地以图像水印为媒介，通过综合运用Harr小波变换、布隆过滤器等先进技术对图像进行预处理，构建出像素点布隆过滤器。同时，利用中国剩余定理的独特性质，对水印信息进行秘密分片处理，实现水印信息的分级、逐条添加，从而能够全面记录并追溯数据的整个流转链路。实验结果显示，该方案不仅具备可逆性，即水印的嵌入与提取过程高效且无损，而且在水印信息编码和嵌入算法的设计上，实现了水印长度的自由设定。此外，该方案在执行效率和鲁棒性方面也展现出了较高的性能，仿真充分验证了其在实际应用中的可行性和优势。

1　相关理论知识

1.1　基于Harr小波变换的图像压缩

Haar小波变换是一种基于Haar小波函数的小波变换[12]，其通过平均和差值操作将原始信号分解为不同频率的成分，得到小波系数，具有计算简单、速度快等优点。

假设图像的4个相邻像素值为$[\alpha,\beta,\chi,\delta]$，基于Haar小波变换的图像压缩的具体步骤如下。

步骤1　将4个像素值两两分组，得到$[\alpha,\beta]$、$[\chi,\delta]$。

步骤2　计算这两组像素值的均值和差值，分别为$[M_0,N_0]$、$[M_1,N_1]$，其中M_i表示均值，$i=0$，1，$M_0=\frac{(\alpha+\beta)}{2}$、$M_1=\frac{(\chi+\delta)}{2}$，$N_i$表示差值，$N_0=\frac{(\alpha-\beta)}{2}$、$N_1=\frac{(\chi-\delta)}{2}$。将这两组值重新组合，成为$[M_0,N_0,M_1,N_1]$。

步骤3　计算$[M_0,M_1]$的均值和差值，进行二级小波系数的交换，其均值和差值表示为$[M_2,N_2]$，其中，$M_2=\frac{M_0+M_1}{2}$、$N_2=\frac{M_0-M_1}{2}$。用$[M_2,N_2]$替换掉步骤2中的$[M_0,M_1]$，就可以得到像素值$[\alpha,\beta,\chi,\delta]$图像最终的二级Haar小波变换系数，为$[M_2,N_0,N_2,N_1]$。

步骤4　根据实际的需求重复步骤2和步骤3，最后提取出由低频系数M_2组成的水印低频图像。

1.2　布隆过滤器

伯顿·霍华德·布隆在1970年提出了布隆过滤器[13]的概念。布隆过滤器是一种概率型数据结构，主要用于判断某个元素是否存在于集合中。

布隆过滤器的主要工作原理：用一个长m（单位为b）的数组表示这个布隆过滤器，数组的每一位元素初始化为0，随后随机选择k个哈希函数，将数据域$S=(y_1,y_2,\cdots,y_n)$中的n个元素分别一一映射到上面的数组中。当需要映射1个元素时，计算该元素的k个哈希函数对应的k个哈希值，然后把数组中对应的位置设置为1，如果哈希值对应的位置已经设置为1，则保留第1次作用的效果。布隆过滤器示意图如图1所示。

要判断元素x是否存在于数据域中，需要计算元素x对应的k个哈希函数的哈希值。如果该元素的所有哈希值对应的位置都为1，就说明元素x存在于数据域中。

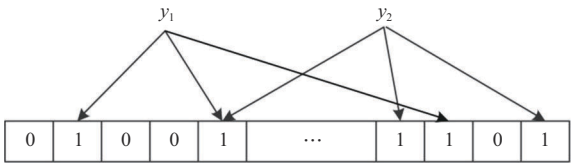

图1 布隆过滤器示意图

当向布隆过滤器中加入1个新的元素时，需要使用k个不同的哈希函数对其进行哈希计算，将这个元素映射到布隆过滤器的k个比特位上，同时将这些比特位设置为1。

在查询某个元素是否存在于布隆过滤器中时，需要利用提前定义的k个哈希函数将其映射到布隆过滤器的k个比特位上。如果这k个比特位对应的值都是1，则说明该元素存在于集合中，只要有一个位置不为1，则可以确定此元素不存在于该集合中。

1.3 基于中国剩余定理的秘密分享方案

秘密分享技术[14]是一种将秘密数据分割成多个部分，并分配给不同参与者的密码学方法。本文主要阐述基于中国剩余定理的秘密分享方案。

首先给出中国剩余定理[15]的性质：给定正整数$m_1, m_2, \cdots, m_n$，当$i \neq j$时，$(m_i, m_j) = 1$，对任意正整数$p_1, p_2, \cdots, p_n$，方程组$x \equiv p_i (\bmod m_i), i = 1, 2, \cdots, n$有解，且当模数$m = m_1 \cdot m_2 \cdot \cdots \cdot m_n$时，有唯一解：$x \equiv \frac{m}{m_1} p_1 x_1 + \frac{m}{m_2} p_2 x_2 + \cdots + \frac{m}{m_n} p_n x_n (\bmod m)$，其中，$\frac{m}{m_i} x_i \equiv 1 (\bmod m_i), i = 1, 2, \cdots, n$。

秘密分享技术是将秘密x分成n个份额$\{x_1, x_2, \cdots, x_n\}$，所有的份额满足任意少于$t$个份额就无法揭示任何关于秘密$x$的信息，而任意不少于$t$个份额就可以恢复原来的秘密$x$，其中$t \leqslant n$，记为$(t, n)$秘密分享方案。下面给出基于中国剩余定理的$(t, n)$秘密分享方案。

令M为需要分享的秘密信息，α、$(\beta_1, \beta_2, \cdots, \beta_n)$为分别两两互素的正整数[16]，$\alpha > M$。

设$\beta_1 < \beta_2 < \cdots < \beta_n$，定义$r$为一个中间计数器，使得$\beta_1 \beta_2 \cdots \beta_r > \alpha \beta_n \beta_{n-1} \cdots \beta_{n-r+2}$。定义$l < \frac{\beta_1 \beta_2 \cdots \beta_r}{\alpha}$，$L = M + l \times \alpha < \beta_1 \beta_2 \cdots \beta_r$。令$L \equiv M_i (\bmod \beta_i)$，$i = 1, 2, \cdots, n$，则$M_i$为秘密信息$M$的秘密份额。

根据中国剩余定理可证明，若已知秘密份额M_i的个数多于或等于r，则可恢复完整秘密M。

2 MIWC介绍

本文提出一种基于中国剩余定理的强鲁棒性多级数据库图像水印嵌入方案MIWC。该方案利用中国剩余定理对经过预处理后的图像水印信息在数据库中进行分级插入，可依次添加多级水印信息，实现对数据整个流转链路的记录和追溯，同时可以保证水印被提取出之后无损地恢复原数据库。MIWC总框架如图2所示。MIWC包含图像水印预处理、图像水印嵌入和图像水印提取共3个阶段。下面分别对各阶段进行介绍。

2.1 图像水印预处理

传统结构化数据水印方案是直接将字符串或二进制比特串作为水印信息嵌入数据中，这种方式可承载的水印信息长度有限，且会引起原数据的较大失真。为了调和嵌入容量与数据失真之间的冲突，力求在嵌入更多身份信息的同时，减少水印的长度，本文使用图像信息来承载要嵌入的水印信息，在水印嵌入数据之前，预先对水印图像进行处理。水印图像的预处理过程包含3个步骤：对图像进行Harr小波变换的图像压缩、将关键像素点的坐标编码输入布隆过滤器、基于中国剩余定理将布隆过滤器值进行秘密分片。

首先定义数据库形式DB:$(P, A_1, A_2, \cdots, A_i, \cdots, A_N)$，其中，$P$为数据库每个元组的主键，$A_i (1 \leqslant i \leqslant N)$为DB的第$i$个属性列，DB中包含$M$个元组，记为$(S_1, S_2, \cdots, S_j, \cdots, S_M)$DB:$(P, A_1, A_2, \cdots, A_i, \cdots, A_N)$，每个元组$S$都有唯一的主键和$N$个属性列。本文利用Harr小波变换技术，对含有水印信息的二维码图像进行了频域层面的分解处理。由于图像中

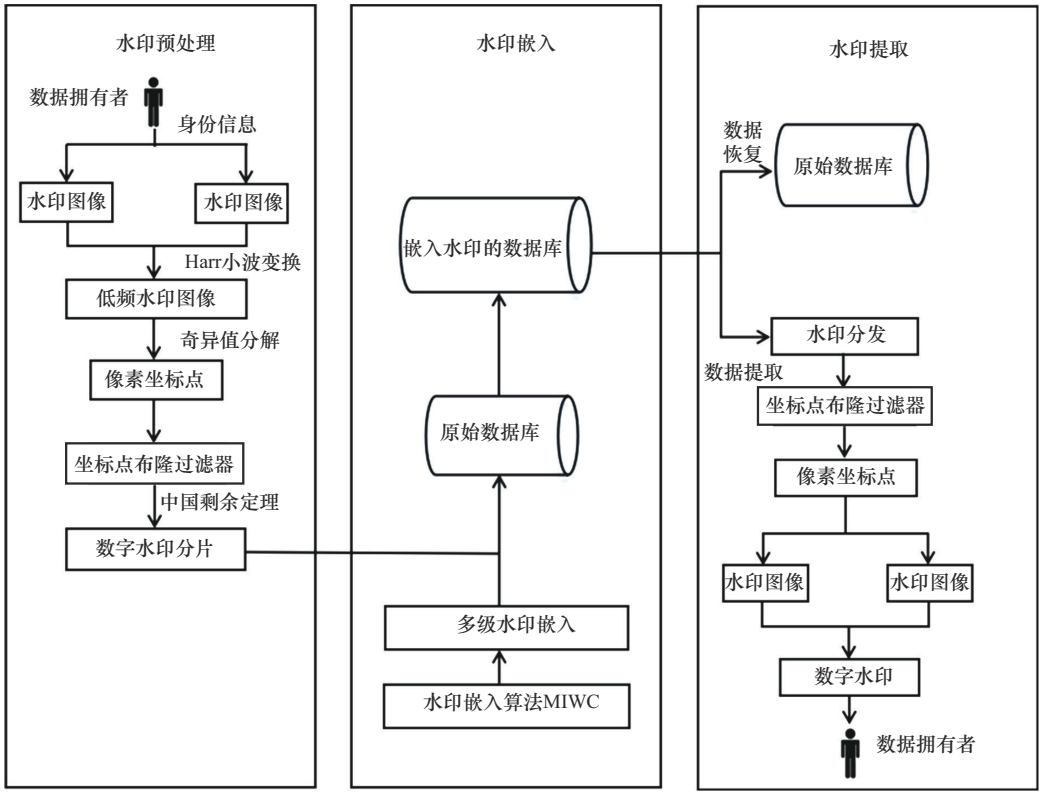

图2　MIWC总框架

的大部分有用信息都集中在低频部分，因此，MIWC特别提取了经过Harr小波变换后得到的低频图像部分，将其作为处理完成的水印图像以便后续应用。具体地，参考第1.1节所述，假设水印图像的4个像素值分别为$\{a_1,a_2,a_3,a_4\}$。随后，方案将这4个像素值逐一代入Harr小波变换的4个关键步骤中。经过一系列变换处理后，成功得到了低频图像，该图像即提取出的水印信息，将用于后续的分析或应用。

接下来，对经过Harr小波压缩后的低频图像进行坐标化处理。拟定低频图像大小为32像素（pm）×32像素（pm），以1 pm为单位建立坐标轴，将低频图像全部坐标化。由于本方案中图像水印的载体为二维码，因此，图像仅有黑、白两种颜色，提取图像中全部黑色点位的坐标（像素值非255），记为坐标集$P=\left\{P_i=\left(x_i,y_i\right),i=1,2,\cdots,n\right\}$。对集合$P$中每一个坐标$P_i$，将其横纵轴坐标合并，得到$m_i=x_i\|y_i$并放入集合$M$中。根据第1.3节所述，构造一个布隆过滤器B，将集合M中的全部元素加入B中，得到像素点布隆过滤器B_m。随后，将低频图像的尺寸插入B_m前两个数组中。尽管布隆过滤器自身具有一定的误识别率，导致B_m相较于原图像具有一定的误差，但是在水印提取过程中，完整的图像水印仍会被完整提取，该误差并不影响水印提取的结果。

由于B_m的数据长度受限，因此，在嵌入数据前需要将B_m进行拆分。如果直接按顺序拆分B_m，那么后续顺序打乱或者丢掉若干分片，将无法恢复B_m，因此，方案必须采用一种鲁棒性更强的数据分片算法。参考第1.3节执行基于中国剩余定理的秘密分片，最终将B_m值分解为一组水印分片W，此时整个方案预处理过程执行完毕。MIWC图像水印预处理流程如图3所示。

图3　MIWC图像水印预处理流程

2.2　图像水印嵌入

得到水印分片$\boldsymbol{W}$之后，接下来需要将这些分片插入数据中，实现水印信息嵌入。假设水印共有N个分片$(W_0,W_1,W_2,\cdots,W_N)$。首先定义结构化数据形式为DB:$(P,A_1,A_2,\cdots,A_i,\cdots,A_{t-1})$，其中，$P$为数据元组的标识符，$A_i(1\leqslant i\leqslant t)$为DB的第$i$个属性列，其中有$S$个数值型属性列。MIWC图像水印嵌入流程如图4所示。

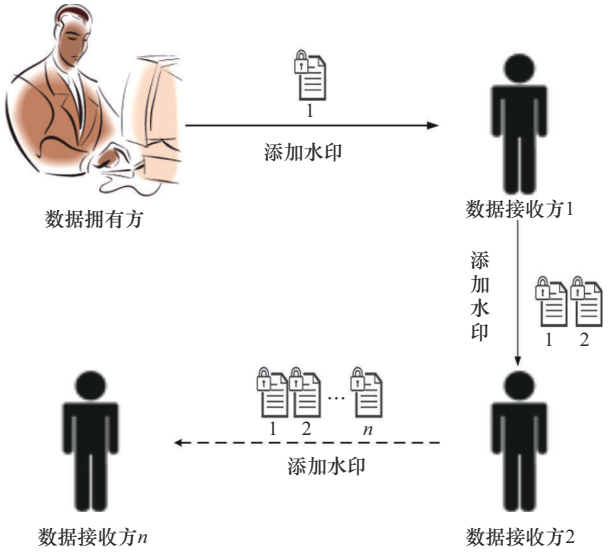

图4　MIWC图像水印嵌入流程

以二级水印嵌入为例，MIWC执行水印嵌入算法步骤如下。

步骤1　数据拥有方确定DB中每一条数据的主键P，计算每一条记录P的基于哈希函数和密钥的消息认证码（hash-based message authentication code，HMAC）值，h=HMAC$_k(P)$，其中，k为数据拥有方的密钥，接着按照h mod MN的值（0,1,2,⋯,MN−1）将逐条数据分为MN组（M为冗余度，用于水印信息纠错），数据库水印分组形式如图5所示。

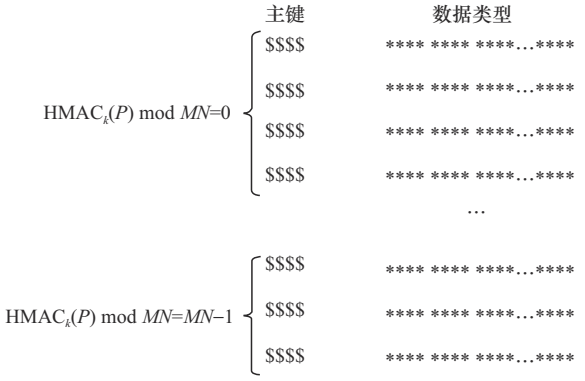

图5　数据库水印分组形式

步骤2　以M=3为例，对于分组$i,i+N,i+2N\left(i\in[0,N-1]\right)$，构造一条伪数据，用于存储水印分片$W_i$，将该条数据标识符$P$的最低位赋值为$i$，其他位随机生成，需要保证最终随机构造的标识符$P'$满足HMAC$_k(P')$ mod MN=i。将第i mod S个数值型属性的最低位赋值为第i个水印分片W_i，对于剩下的S−1个数值型属性，将第j个数值型属性的最低位赋值为i mod j，其余字段按照原数据的格式随机生成。

步骤 3　当数据接收方 1 需要将数据转发给数据接收方 2 时，重复上述步骤向数据中添加接收方 2 的水印，唯一不同之处是在步骤 1 中计算 $h=\mathrm{HMAC}_{k_1}(P)$ 时所用的密钥 k_1 为接收方 1 的密钥。

在二级水印嵌入的过程中，对数据库的分组以及每一水印分片的嵌入均为独立的，每一步骤互相均不受影响。因此，方案可以从二级水印嵌入扩展至多级水印嵌入，即当数据参与方数目为 n 时，只需要依次重复上述步骤 $n-1$ 次，便能够实现多级水印嵌入。

以二级图像水印嵌入为例，给出 MIWC 图像水印嵌入步骤的伪代码如下。

输入　原始数据库 DB，经过预处理的图像水印 watermark_info；

输出　嵌入多级图像水印的数据库 DB*。

```
function encode_watermark(DB, watermark_info, N, L, M, S, k):
    encoded_watermark = encode(watermark_ info)
    watermark_shares=chinese_remainder_theorem_ share(encoded_watermark, N, L)
    if DB has_primary_key(data): P = get_primary_key(data)
    else: P = construct_unique_attribute_set(data)
    HMAC_values = calculate_hmac(P, k)
      for i from 0 to MN-1:
        group_index = hmac_values[i] % (MN)
      if (i % N) == 0: pseudo_data= construct_pseudo_data(data, i, M, S, watermark_shares[i // N])
    return processed_data
function construct_pseudo_data(data, i, M, S, watermark_share):
    pseudo_record = create_new_record()
    pseudo_record.P_last_digit = i
    for j from 0 to S-1:
      if j == (i % S):
        pseudo_record. numeric_attribute[j] =set_lsb (pseudo_record. numeric_attribute[j], watermark_share)
      else: pseudo_record. numeric_attribute[j] = set_lsb(pseudo_record. numeric_attribute[j], i % j)
      pseudo_record. other_fields = generate_random_fields(data.T)
      return pseudo_record
data_with_watermark1 = encode_watermark (DO1, data, watermark_info1, N, L, M, S, k)
data_with_watermark2=encode_watermark (DO2,data_with_watermark1,
    watermark_info2, N, L, M, S, k2)
insert data_with_watermark1， data_with_watermark2 into DB
return DB*
```

2.3　图像水印提取

在遇到数据权属纠纷或第三方数据泄露事件时，可以通过提取数据中的水印信息来进行权属证明和责任方认定。MIWC 图像水印提取步骤如下。

步骤 1　数据拥有方设置 N 个数组 $R_0[],R_1[],R_2[],\cdots,R_{N-1}[]$，用自己的密钥 k 计算 $\mathrm{HMAC}_k(P)\bmod MN$ 对持有水印的数据库 DB*进行分组，对于分组 $i,i+N,i+2N(i\in[0,N-1])$，找出满足以下两个条件的数据行。

（1）标识符最低位的值与分组编号相同。

（2）对于分组的第 j 个数值型属性，如果 $j\neq i(i+N,i+2N)$，那么属性的最低位等于 $i\bmod j$。

如果数据行满足上述条件，则取出第 $i\bmod S$ 个数值型属性的最低 L 位，放入数组 $R_i[]$。

步骤 2　对于数组 $R_0[],R_1[],R_2[],\cdots,R_{N-1}[]$，取

出每个数组中出现频次最高的值，构成N个水印分片，然后执行分片重组算法，得到水印值BL。

步骤3 提取出的水印值BL为布隆过滤器的值，需要将其恢复为水印图像。首先提取BL前两个字节得到水印图像的尺寸信息，按该尺寸初始化为一张白色背景图片，然后遍历尺寸坐标的每一个坐标值，判断其能否通过布隆过滤器的筛选。如果能通过，则将该坐标对应的像素灰度置为“黑”（即非255的像素值），最终恢复水印图像。

值得注意的是，即便还原的黑白图像出现了一定程度的失真，只要这种失真保持在一定的范围内，方案依然能够扫描出身份信息。由于经过分片后的图像水印被依次完整提取，数据库信息得以被完全复原，因此，MIWC方案具备可逆性。该特性充分展示了本方案在安全性与鲁棒性方面的优势。

下面给出MIWC图像水印提取步骤的伪代码。

输入 持有多级图像水印的数据库DB*，数据拥有者密钥k，布隆过滤器BL；

输出 恢复后的图像IMAGE。

```
function group_and_extract_watermark(DB*, k, N, L):
    watermark_bits = []
    group_size = len(DB*)
    for i in range(N):
        group = DB[i * group_size: (i + 1) * group_size]
        group_bits = []
    for row in group:
        dentifier_last_bit = row[0] & 1
        group_number_last_bit = i & 1
    condition_met = (identifier_last_bit == group_number_last_bit) )
    if condition_met:
        bit_to_extract = row[k] & ((1 << L) - 1)
        group_bits.append(bit_to_extract)
        watermark_bits.append(group_bits)
        watermark_fragments = []
    for bits in watermark_bits:
        most_frequent_bit = find_most_frequent(bits)
        watermark_fragments. append(most_frequent_bit)
    BL = reassemble_watermark(watermark_fragments)
    image_size = extract_image_size(BL)
    image = create_blank_image(image_size)
    for x in range(image_size.width):
        for y in range(image_size.height):
            coordinate_hash = hash_coordinate(x, y)
            if bloom_filter_test(BL, coordinate_hash):
                image.set_pixel(x, y, BLACK)
    return IMAGE
```

3 实验测试

3.1 系统实施架构

本文MIWC的仿真实验采用中国移动某省公司能开平台日志数据作为测试数据，承载水印信息的黑白图像集来自麻省理工学院计算机科学和人工智能实验室共同创建的大型数据集Lableme，图像尺寸为256像素×256像素，存储数据的数据库为MySQL 8.0。MIWC测试数据集示例如图6所示，测试集中Lableme中待嵌入原始图像示例如图7所示。

本文MIWC的实验测试是一个综合性的过程，旨在评估该方案在实际应用中的可行性，主要评估指标包括方案的水印嵌入效率、方案的鲁棒性，并在功能和性能两个角度将本方案与当前业界的数据

city_id	province_id	starttimehh	service_id	subservice_id	app_server_ip
0851	851	2023103105	01	00002	180. 76. 76. 76
0851	851	2023103105	01	00002	211. 139. 5. 27
0851	851	2023103105	01	00002	211. 139. 5. 28
0851	851	2023103105	01	00002	2409:806A:2010:0000:0000:0000:0000:0001
0851	851	2023103105	01	00002	2409:806A:2010:0000:0000:0000:0000:0002
0851	851	2023103105	01	00005	1. 0. 0. 1
0851	851	2023103105	01	00005	1. 1. 1. 1
0851	851	2023103105	01	00005	1. 13. 155. 223
0851	851	2023103105	01	00005	1. 189. 232. 50
0851	851	2023103105	01	00005	10. 100. 143. 197
0851	851	2023103105	01	00005	10. 117. 0. 159
0851	851	2023103105	01	00005	10. 140. 0. 71
0851	851	2023103105	01	00005	10. 48. 168. 147
0851	851	2023103105	01	00005	10. 88. 11. 125
0851	851	2023103105	01	00005	101. 133. 143. 203
0851	851	2023103105	01	00005	101. 133. 229. 104
0851	851	2023103105	01	00005	101. 226. 142. 111
0851	851	2023103105	01	00005	101. 226. 144. 238
0851	851	2023103105	01	00005	101. 35. 212. 35
0851	851	2023103105	01	00005	101. 35. 76. 130

city_id	province_i	starttime	service_i	subservic	app_serve	ultraffic	dltraffic	httpultraf
851	851	2. 023E+09	1	2	180. 76. 76.	202	146	
851	851	2. 023E+09	1	2	211. 139. 5.	76770	18894	
851	851	2. 023E+09	1	2	211. 139. 5.	912	800	
851	851	2. 023E+09	1	2	2409:806A	36721	27813	
851	851	2. 023E+09	1	2	2409:806A	1367	1300	
851	851	2. 023E+09	1	5	1. 0. 0. 1	508	387	
851	851	2. 023E+09	1	5	1. 1. 1. 1	0	1083	
851	851	2. 023E+09	1	5	1. 13. 155. 2	1144	2423	2423
851	851	2. 023E+09	1	5	1. 189. 232.	521	589	589
851	851	2. 023E+09	1	5	10. 100. 143	0	79	
851	851	2. 023E+09	1	5	10. 117. 0. 1	0	768	
851	851	2. 023E+09	1	5	10. 140. 0. 7	0	172	
851	851	2. 023E+09	1	5	10. 48. 168.	0	1794	
851	851	2. 023E+09	1	5	10. 88. 11. 1	0	2233	
851	851	2. 023E+09	1	5	101. 133. 14	1066006	34770	34770
851	851	2. 023E+09	1	5	101. 133. 22	1603	1581	
851	851	2. 023E+09	1	5	101. 226. 14	34119	67957	67957
851	851	2. 023E+09	1	5	101. 226. 14	5280	18166	18166
851	851	2. 023E+09	1	5	101. 35. 212	3153	5218	5218
851	851	2. 023E+09	1	5	101. 35. 76.	1897	5030	5030
851	851	2. 023E+09	1	5	101. 35. 76.	306	554	554
851	851	2. 023E+09	1	5	103. 106. 2(	107650	23995	23995

图6　MIWC测试数据集示例

库水印方案进行对比。下面给出方案需要分析的实验指标、实验结果以及方案对比。

图7　测试集中Lableme中待嵌入原始图像示例

3.2　实验指标

方案的实验指标包含方案的功能指标与性能指标两个部分，共6项指标。下面分别对这6项待测试的实验指标进行详细介绍。

（1）功能指标

水印载体指的是待嵌入至数据库水印所支持的类型；可逆性指方案是否支持对嵌入的数据库水印进行逆向提取；多级插入指嵌入的水印序列是否支持多级水印插入。

对于功能指标，将MIWC与现有数据库水印方案[17-20]进行对比，进一步证明本文方案的全面性与实用性。MIWC同各方案的功能对比结果见表1。

由表3可知，MIWC相比于现有数据库水印方案，不仅仅支持多种类型的水印载体，且具备可逆性，最重要的是，MIWC支持水印的多级插入，即新插入的水印不会覆盖之前的水印，在功能上具备一定优势。

（2）性能指标

水印鲁棒性指删除部分数据，仍能正常提取水印的概率，正确提取水印的概率越大，证明水印自身的鲁棒性越强；水印嵌入效率指评估水印嵌入至数据库中的效率，通常以嵌入时间长短来评价该指标，时间越短，证明方案在分级水印嵌入数据库的过程效率越高，实用性越强；水印提取效率指评估水印从数据库中提取的效率，其评估指标等同于水印嵌入效率。

3.3　实验结果

本文首先给出MIWC在图像水印嵌入数据库前后的变化，嵌入水印前后数据库对比如图8所示。

表1　MIWC同各方案的功能对比结果

方案	水印载体	可逆性	支持多级插入	关键技术
MIWC	字符串/图片	√	√	Harr小波+中国剩余定理
文献[17]	字符串	×	×	中国剩余定理
文献[18]	字符串/图片	√	×	动态差分扩展
文献[19]	字符串/图片	√	×	Harr小波+动态差分扩展
文献[20]	字符串/图片	√	×	双层插值嵌入

由图8可以看出，由于图像已被预处理分片，因此整体数据库形式没有变化，仅有脱敏数据的改变。下面分别对方案的3项性能指标进行测试。

（1）鲁棒性测试

在对方案的鲁棒性进行测试时，本文主要考虑删除攻击，即删掉原数据的部分内容，测试能正确提取水印值的概率。在实际的水印攻击中，攻击者通过删除数据库中的元素以破坏水印，当删除的水印分片超过能接受的阈值范围时，便可产生攻击效果，导致水印信息无法被正确恢复。水印鲁棒性实验及方案对比如图9所示。

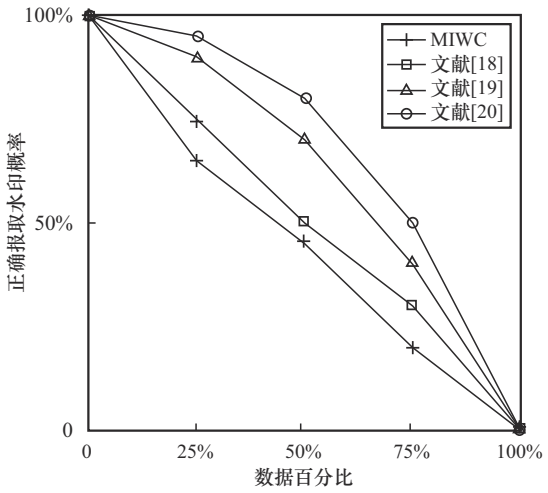

图9　水印鲁棒性实验及方案对比

由图9可知，在面临数据删除攻击时，本文所提方案具备较强的鲁棒性。

（2）水印嵌入测试

由于本文方案为分级水印添加，现有方案并没有类似的处理方式，因此在测试过程中只考虑一级嵌入时的效率，并与现有数据库水印方案进行对比。单极水印嵌入实验及方案对比如图10所示。

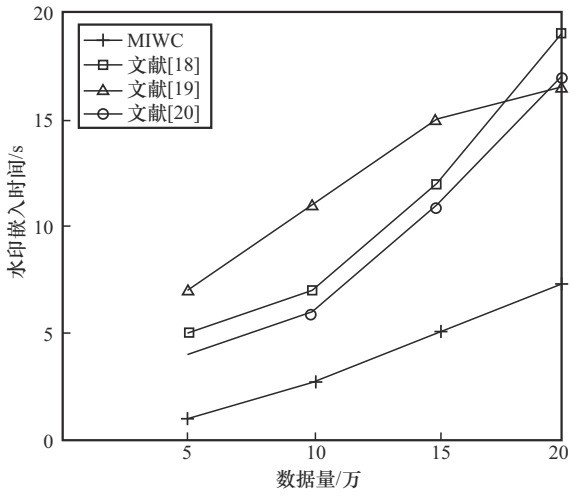

图10　单极水印嵌入实验及方案对比

由图10可知，本文方案在一级水印嵌入的过程中，相比于传统的单极图像数据库水印方案，具备优势。由于方案支持多级嵌入，因此在多级水印累加嵌入的过程中，MIWC的嵌入效率优势将会更明显。

（3）水印提取测试

对不同大小的数据集进行水印提取性能测试，水印提取实验及方案对比如图11所示。

由图11可知，MIWC在水印提取时，相比于数据库水印方案，效率明显更优，因此，具备更高的应用价值。

city_id	province_	starttime	service_i	subservic	app_servi	ultraffic	dltraffic	httpultra
851	851	2.023E+09	1	2	180.76.76.	202	146	
851	851	2.023E+09	1	2	211.139.5.	76770	18894	
851	851	2.023E+09	1	2	211.139.5.	912	800	
851	851	2.023E+09	1	2	2409:806A	36721	27813	
851	851	2.023E+09	1	2	2409:806A	1367	1300	
851	851	2.023E+09	1	5	1.0.0.1	508	387	
851	851	2.023E+09	1	5	1.1.1.1	0	1083	
851	851	2.023E+09	1	5	1.13.155.	1144	2423	2423
851	851	2.023E+09	1	5	1.189.232.	521	589	589
851	851	2.023E+09	1	5	10.100.14	0	79	
851	851	2.023E+09	1	5	10.117.0.	0	768	
851	851	2.023E+09	1	5	10.140.0.	0	172	
851	851	2.023E+09	1	5	10.48.168.	0	1794	
851	851	2.023E+09	1	5	10.88.11.	0	2233	
851	851	2.023E+09	1	5	101.133.14	1066006	34770	34770
851	851	2.023E+09	1	5	101.133.22	1603	1581	
851	851	2.023E+09	1	5	101.226.14	34119	67957	67957
851	851	2.023E+09	1	5	101.226.14	5280	18166	18166
851	851	2.023E+09	1	5	101.35.212	3153	5218	5218
851	851	2.023E+09	1	5	101.35.76.	1897	5030	5030
851	851	2.023E+09	1	5	101.35.76.	306	554	554
851	851	2.023E+09	1	5	103.106.20	107650	23995	23995

（a）嵌入水印前

city_id	province_	starttime	service_i	subservic	app_servi	ultraffic	dltraffic	httpultra
851	851	2.023E+09	1	2	180.76.76.	202	146	
851	851	2.023E+09	1	2	211.139.5.	76770	18894	
851	851	2.023E+09	1	2	211.139.5.	912	800	
851	851	2.023E+09	1	2	2409:806A	36721	27813	
851	851	2.023E+09	1	2	2409:806A	1367	1300	
851	851	2.023E+09	1	5	1.0.0.1	508	387	
851	851	2.023E+09	1	5	1.1.1.1	0	1083	
851	851	2.023E+09	1	5	1.13.155.	3458	2423	2423
851	851	2.023E+09	1	5	1.189.232.	521	589	589
851	851	2.023E+09	1	5	10.100.14	0	79	
851	851	2.023E+09	1	5	10.117.0.	0	768	
851	851	2.023E+09	1	5	10.140.0.	0	172	
851	851	2.023E+09	1	5	10.48.168.	0	1794	
851	851	2.023E+09	1	5	10.88.11.	0	2233	
851	851	2.023E+09	1	5	101.133.14	1066006	34770	34770
851	851	2.023E+09	1	5	101.133.22	1603	1581	
851	851	2.023E+09	1	5	101.226.14	34119	67957	67957
851	851	2.023E+09	1	5	101.226.14	5280	18166	18166
851	851	2.023E+09	1	5	101.35.212	3153	5218	5218
851	851	2.023E+09	1	5	101.35.76.	1897	5030	5030
851	851	2.023E+09	1	5	101.35.76.	306	554	554
851	851	2.023E+09	1	5	103.106.20	107650	23995	23995

（b）嵌入水印后

图8　嵌入水印前后数据库对比

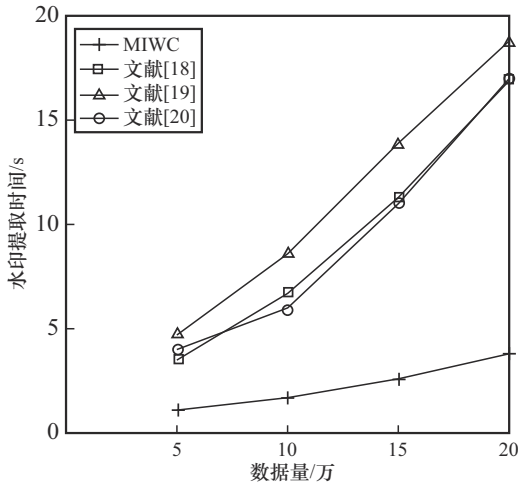

图 11　水印提取实验及方案对比

3.4　方案应用情况

本文结构化数据水印方案 MIWC 已自研实现，并在中国移动贵州省公司落地试点，主要针对“东数西算”和数据流通共享场景，解决数据跨域跨主体流转所面临的权属问题和泄露之后的追踪溯源问题，MIWC 应用架构如图 12 所示。相对于传统的本地处理数据的模式，数据外包模式在数据的上传、分发、回传等数据流转的过程中，存在数据泄露或滥用的风险，本文方案则在数据跨域节点嵌入水印，实现数据确权和流转追踪溯源，降低可能的数据安全风险。

4　结束语

随着数据业务的发展，企业对于数据的需求越来越大，自身数据往往难以满足高速增长的业务需求，因此，数据流通成为发展的必然趋势。由于重要数据、个人信息数据等存储量较大、分类分级困难，明确数据权属和数据发生泄露事件时的责任归属，是保障数据安全的重要手段，也是确保数据能够安全可靠流通的前提。数字水印技术是解决这一问题的有效手段，但传统水印技术在应用于结构化数据时存在诸多缺陷。本文所提方案自主设计了新的水印信息编码和水印嵌入算法，提出了一种针对结构化数据的多级水印方案 MIWC，该方案在功能性、执行效率、鲁棒性方面均有明显优势。

在后续的工作推进中，基于水印的数据确权仍然存在需要继续研究和完善的内容，水印技术的执行效率、鲁棒性、易用性仍存在优化的空间。除此之外，可以将水印技术与其他多种技术融合来实现数据确权，如基于区块链技术进行数据操作的上链存证溯源，一旦出现权属纠纷可以通过链上记录来查证。区块链与数据水印也可以互相补齐对方的短板，如数据水印可以解决链下的数据权属声明问题，区块链可以优化水印导致的数据失真和易用性的问题。数据水印技术也可以与数据分类分级联动，从而可针对不同级别的数据提供不同的确权措施。在现网的环境下，随着 AI 大模型、云计算、算力网络等技术的发展，基于数据水印的数据确权技术将面临更广阔的应用场景。

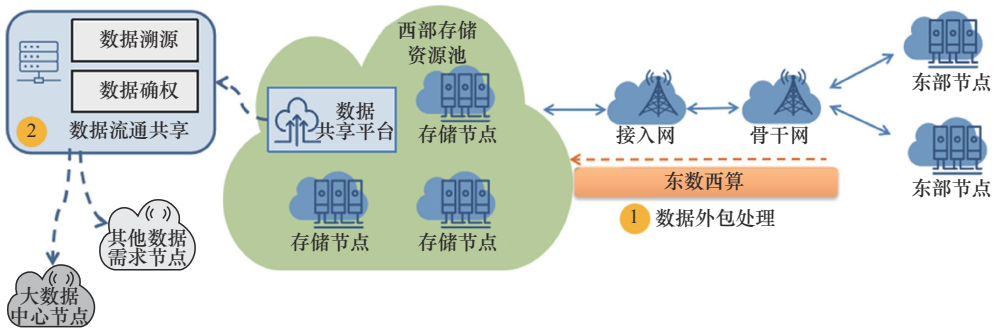

图 12　MIWC 应用架构

参考文献：

[1] 朱勤, 于守健, 乐嘉锦. 数据库水印研究与进展[J]. 计算机工程与应用, 2006, 42(29): 198-201.
ZHU Q, YU S J, LE J J. Research work and progress in database watermarking[J]. Computer Engineering and Applications, 2006, 42(29): 198-201.

[2] 刘珺, 罗守山, 吴秋新, 等. 基于中国剩余定理的数字水印分存技术[J]. 北京邮电大学学报, 2002, 25(1): 17-21.
LIU J，LUO S S，WU Q X, et al. Digital watermark sharing based on Chinese remainder theorem[J]. Journal of Beijing University of Posts and Telecommunications, 2002, 25(1): 17-21.

[3] 曾皓炜, 白恩健, 吴赟. 基于中国剩余定理和混沌系统的高低位图像分治加密[J]. 图像与信号处理, 2023, 12(2): 116-127.
ZENG H W, BAI E J, WU Y. Divide-and-conquer encryption of high and low bit images based on Chinese remainder theorem and chaotic system[J]. Journal of Image and Signal Processing, 2023, 12(2): 116-127.

[4] 陶伟成, 李智勇, 李厚甫. 基于极角扩展的可逆盲数据库水印算法[J]. 计算机工程, 2010, 36(22): 155-157.
TAO W C, LI Z Y, LI H F. Reversible and blind database watermark algorithm based on polar angle expansion[J]. Computer Engineering, 2010, 36(22): 155-157.

[5] 崔新春, 贺洁, 秦小麟. 基于盲源分离的多重音频数据库水印算法[J]. 电子学报, 2012, 40(1):80-85.
CUI X C, HE J, QIN X L. A multiple audio watermarking for database based on blind source separation[J]. Acta Electronica Sinica, 2012, 40(1):80-85.

[6] DELAIGLE J F, DE VLEESCHOUWER C, MACQ B M M, et al. Digital watermarking[C]//Proceedings of the International Society for Optical Engineering(SPIE), Piscataway: IEEE Press, 1996: 99-110.

[7] KHALLAF F, EL-SHAFAI W, EL-RABAIE E S M, et al. A novel hybrid cryptosystem based on DQFrFT watermarking and 3D-CLM encryption for healthcare services[J]. Frontiers of Information Technology & Electronic Engineering, 2023, 24(7): 1045-1061.

[8] ZHANG Z W, WANG H, WANG G S, et al. Hide and track: towards blind video watermarking network in frequency domain[J]. Neurocomputing, 2024, 579: 127435.

[9] CHEN X, CHEN P, HE Y, et al. A self-resilience digital image watermark based on relational database[C]//Proceedings of International Symposium on Knowledge Acquisition & Modeling. Piscataway: IEEE Press, 2008: 698-702.

[10] CHANG C, NGUYEN T, LIN C C. A reversible database watermark scheme for textual and numerical datasets[C]//Proceeding of the 2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Piscataway: IEEE Press, 2021: 208-212.

[11] XIAN H, LI J, LU X. Tracing mobile database with complex watermark[J]. International Journal of Database Theory and Application, 2015, 8: 23-28.

[12] MONTAZER M, KHODABIN M, EZZATI R, et al. Numerical solution of stochastic Volterra integral equations based on uniform Haar wavelets by using direct method[J]. Asian-European Journal of Mathematics, 2023, 16(10): 16.

[13] HADIAN DEHKORDI M, MASHHADI S, FARAHI S T, et al. Changeable essential threshold secret image sharing scheme with verifiability using Bloom filter[J]. Multimedia Tools and Applications, 2024, 83(20): 58745-58781.

[14] IQBAL S, RAUF A, MAHFOOZ S, et al. Self-constructing fragile watermark algorithm for relational database integrity proof[J]. World Applied Sciences Journal, 2012, 19(9): 1273-1277.

[15] CHOWDHURY A A, KESSERWANI G, ROUGÉ C, et al. GPU-parallelisation of Haar wavelet-based grid resolution adaptation for fast finite volume modelling: application to shallow water flows[J]. Journal of Hydroinformatics, 2023, 25(4): 1210-1234.

[16] 丁妍, 姚爱琴, 张智尊. 基于三次B样条小波变换和Canny算法的火焰边缘检测算法[J]. 无损检测, 2023, 45(4): 43-47.
DING Y, YAO A Q, ZHANG Z Z. Flame edge detection algorithm based on cubic B-spline wavelet transform and Canny algorithm[J]. Nondestructive Testing Technologying, 2023, 45(4): 43-47.

[17] 张桂芳, 孙星明, 肖湘蓉, 等. 基于中国剩余定理的数据库水印技术[J]. 计算机工程与应用, 2006, 42(7): 135-136, 173.
ZHANG G F, SUN X M, XIAO X R, et al. Database watermarking based on Chinese remainder theorem[J]. Computer Engineering and Applications, 2006, 42(7): 135-136, 173.

[18] PATEL R. Relational database watermarking clustering based data-mining using K-Means approach[J]. International Journal for Scientific Research & Development, 2015, 3(3): 1491-1495.

[19] 姚绍华, 陈青, 陈祥. 一种强鲁棒性的新型彩色图像数字水印技术[J]. 数据通信, 2017(4): 27-30.
YAO S H, CHEN Q, CHEN X. A novel color image digital watermarking technique with strong robustness [J]. Data Communications, 2017(4): 27-30.

[20] 汪天琦, 张迎周, 邸云龙, 等. 基于动态差分扩展的强鲁棒数

据库水印算法研究[J]. 网络与信息安全学报, 2023, 9(5): 150-165.
WANG T Q, ZHANG Y Z, DI Y L, et al. Research on strong robustness watermarking algorithm based on dynamic difference expansion[J]. Chinese Journal of Network and Information Security, 2023, 9(5): 150-165.

[作者简介]

耿慧拯（1988– ），男，中国移动通信有限公司研究院安全所高级工程师，主要研究方向为数字水印、隐私计算以及数据安全。

郭斯栩（1996– ），男，中国移动通信有限公司研究院安全所工程师，主要研究方向为数字水印、密码学与可搜索加密。

刘颖卿（1989– ），女，中国移动通信有限公司研究院安全所工程师，主要研究方向为数字水印、密码学与可搜索加密。

粟栗（1981– ），男，博士，中国移动通信有限公司研究院安全所副所长、正高级工程师，主要研究方向为移动通信网络安全和数据安全。

何申（1980– ），男，博士，中国移动通信有限公司研究院安全所所长、正高级工程师，主要研究方向为网络安全、物联网安全、移动互联网安全、可信计算等。