

基于隐写技术的气象数据安全保护研究

郭 聪,马 强,张子剑
(国家气象信息中心,北京 100081)

摘 要:随着科技水平的不断提高,各行各业对气象数据的需求越来越大。技术发展给气象数据共享带来便利的同时,也给气象数据的非法泄露、篡改和滥用的责任追查带来困难。目前,气象数据的安全保护措施还比较匮乏,为了有效解决这一问题,提出了一种高效的可行性方案——INVIEN。该方案能够根据不同的气象数据类型,自适应地采用不同的隐写嵌入算法,将授权用户的用户信息以隐写的方式嵌入原始数据文件中。该方案不仅可以有效解决数据共享后该数据的原授权用户信息不易追查的问题,还可以为数据所有者版权追溯提供必要的技术支撑依据。经过反复测试和试验,该方案可以应用于 TXT 格式和 PNG 格式的气象数据中。通过多次实验分析发现,该方案处理批量数据非常高效、隐蔽。

关键词:隐写技术;气象;数据;信息安全

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2019)08-0119-06

doi:10.3969/j.issn.1673-629X.2019.08.023

Research on Invisible Encryption Algorithm of Meteorological Data Based on Digital Watermarking

GUO Cong, MA Qiang, ZHANG Zi-jian
(National Meteorological Information Center, Beijing 100081, China)

Abstract: With the continuous improvement of science and technology, the demand for meteorological data from all walks of life is increasing. While the technical development brings convenience to the sharing of meteorological data, it also brings difficulty to the responsibility tracing of illegal leakage, tampering and abuse of meteorological data. At present, meteorological data security protection measures are still relatively scarce. In order to effectively solve this problem, an efficient and feasible scheme – INVIEN is proposed. According to different weather data, we use different digital watermarking encryption algorithms to add authorized usage information of data in different locations. Through the decryption and analysis of digital watermark, we can track the weather data of authorized users. This scheme can not only effectively solve the problem that it is difficult to trace the original authorized user information of the data after data sharing, but also provide necessary technical support for the copyright tracing of the data owner. The experimental model is applied to the weather data in TXT and PNG formats. It is found that this scheme is very efficient and covert in processing batch data.

Key words: digital watermark; meteorology; data; information security

0 引 言

随着技术的不断进步,信息媒体的数字化为信息的提取和存储带来了极大便利。数据的交换和传输简单给共享数据的信息安全防护、知识产权保护和数据认证取证带来困难,亟需采取有效措施予以解决。气象数据蕴含丰富的应用和研究价值,同时气象服务需求不断增加。气象部门响应政府号召,采用多种方式对外提供气象资源和数据服务,但笔者发现,当数据共享给授权者后,无法及时采取有效的技术手段阻止非授权用户的非法数据滥用、泄露等问题。这些给数据

的管理和安全防护带来极大挑战。

信息安全随着技术的不断发展,受到越来越多研究者的青睐,数字水印已成为近年来信息安全领域关注的热点^[1-6]。数字水印中的隐写技术是一种可以将秘密信息在不损坏载体质量的情况下,以用户不知情的形式嵌入到数字媒介,从而为数据提供安全标记来保护数据安全的技术。由于其隐蔽不易被发现,信息可以在开放的环境中安全传输而不易被影响。目前,隐写技术的经典算法——空域算法^[7],应用十分广泛。它通过将必要信息嵌入到随机选择的图像点中最不重

收稿日期:2018-07-25

修回日期:2018-11-27

网络出版时间:2019-03-27

基金项目:国家重点研发计划(2016YFB0800301);国家气象信息中心第六届青年科技基金课题(NMICQJ201701)

作者简介:郭 聪(1982-),女,博士研究生,高级工程师,从事网络与信息安全、系统开发与设计相关领域课题的研究。

网络出版地址:<http://kns.cnki.net/kcms/detail/61.1450.TP.20190327.1620.004.html>

要的像素位 (least significant bits, LSB) 上,来保证嵌入的数字标记的不可见,但研究者通过分析发现,该方案技术鲁棒性不高。经典变换域算法应用也十分广泛,主要通过扩展频谱通信技术 Patchwork 算法,对像素亮度进行调整从而加入必要用户信息,从而保护数据安全。这两类算法虽然较易实现,但能够添加的信息十分有限,且没有将图像格式和文本格式同时考虑保护的尝试。根据笔者前期调研分析,气象数据的数据存量,数据种类繁多,文本类数据和图片类数据又尤其重要,因此需要一种相应的数据隐写保护技术,以保护气象数据的安全。

基于上述考虑,提出了一种针对不同类型气象数据的有效信息隐写技术添加方案——INVIEN。该方案可针对不同气象数据的特点,在数据的不同位置添加隐形的数据授权使用信息。当数据共享服务后,可通过对数据中数字水印^[8]的鉴定和分析来核实数据所有者信息,从而为气象数据的版权保护和鉴别提供依据,有利于保护气象数据的数据安全。

文中主要贡献如下:提出了一种新颖的隐写方案,该方案复用性强,适用于大数据量数据的使用;基于密码学的数字签名技术,能够对嵌入信息提供进一步的数据完整性保护;增加了对嵌入信息完整性的自定义校验 CRC 技术^[9-10],能够有效保护数据传输过程中的

数据安全。

据笔者了解,该方案是第一个针对批量气象数据进行数字标记添加的行业用户。通过对批量数据的集中多维度系统测试,结果表明该方案能够对不同类型的数据进行有效隐写水印添加,并能协助鉴别已共享数据的数据所有者权益。

1 加密系统概述

本节主要介绍系统模型,该系统模型基于数据隐写数字水印不可见加密方法,构建并设计了文中方案。

1.1 系统模型

文中设计的系统模型 INVIEN 包含两个主体:用户和服务器。该模型主要针对气象数据中最常见的两种数据格式 PNG 及 TXT 进行保护。用户需要将加密的数据提供给服务器,服务器可以批量完成对不同格式的用户信息的隐写添加。

加密方案可表示为:

$$\pi = (QueryEn, QueryDe)$$

$Data \leftarrow QueryEn$:提交加密查询请求算法,服务器返回加密后的数据结果 Data。

$K \leftarrow QueryDe$:提交解密查询请求算法,服务器返回插入数据中的水印信息 K。

系统模型如图 1 所示。

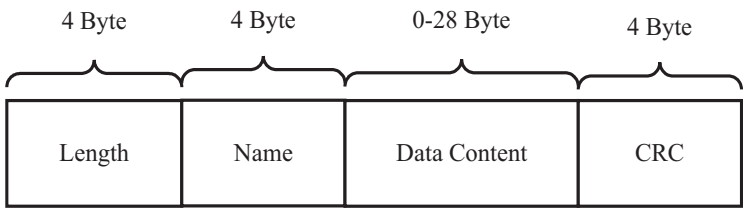


图 1 系统模型

1.1.1 用户

用户,即数据属主,负责向系统提出请求并提供需要加密的数据或请求对已加密数据进行解密,并接收从系统返回的结果。

1.1.2 服务器

接受用户的加密请求或解密请求。对用户提交的批量文件,根据不同类型采用不同模块接收请求。对加密请求返回加密后的数据,解密请求返回加密水印中的信息。

1.2 INVIEN 的构建

INVIEN 系统由两大部分构成:对 PNG 格式的加密;对 TXT 格式的加密。其中用到了多项技术,下面对系统中使用到的技术进行简单介绍。

chunk 的构建描述如下:

这部分是关键,这里单独着重介绍。

针对插入的信息,设计了一个简单的加密消息结

构 chunk,约定消息长度 40 Byte,加密内容在 28 Byte 以内。28 Byte 能满足一般需求,如果需要扩大,可以后续修改。值得注意的是,还在结构中加入了 CRC 校验^[9]防止错误出现。

在此基础上,在 chunk 里基于 MD5 和 RSA 的签名结构对消息进行保护。加密过程如下:

首先生成公钥、私钥并保存,长度均为 512 位。

(PK,SK)

用户信息定义为 userinfo,使用 MD5 对 userinfo 摘要,摘要信息定义为 md5_data。

$$md5_data = md5(userinfo)$$

签名用 RSA 对 md5_data 加密,得到签名 sign_ta。

$$sign_data = RSA(md5_data)$$

最后将 userinfo 和 sign_data 进行拼接,得到 chunk_data,最后包装为 chunk 写入文件。

$$chunk = length + chunkname + CRC32(chunk_$$

`data(userinfo + sign_data) + chunkname) +`
`(chunk_data(userinfo + sign_data) + chunkname)`
加密消息结构如图 2 所示。

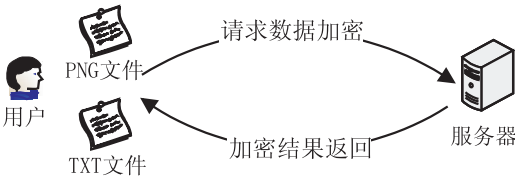


图 2 加密消息结构

解密过程相对,将加密过程倒序执行即可。先拆解 `chunk_data` 得到 `userinfo`,再用 `md5` 得到摘要,然后用 `RSA` 对 `sign_data` 进行解密,并与前面摘要比对,相同则验证通过,获得用户信息。

1.3 设计目标

设计目标是实现一个加密系统,能够对 PNG 及 TXT 格式的批量数据进行加密。加密内容可以自定义,并且对加密数据进行签名,能够进一步保护数据的安全。如果发生数据泄露等问题,可以通过鉴别数据的信息进行追踪溯源。

2 加密算法的思路及构建

2.1 PNG 格式的加密实现

通过对 PNG 文件进行解析,增加包含加密信息的文件块程序模块。这里提出了两种加密思路,其一是文件头加密技术^[11-12],另一种是像素点加工加密。

根据已知的 PNG 文件格式,可知 PNG 文件头位置总是由位固定的字节来描述的。其中第一个字节为了避免某些软件将 PNG 文件误认为文本文件来处理,值 `0x89` 超出了 ASCII 字符的范围用来标识。文件中剩余的部分由 3 个以上的 PNG 的数据块 (`chunk`) 按照特定的顺序组成,因此,一个标准的 PNG 文件结构由文件标识及众多 `chunk` 构成。

PNG 定义了两种类型的数据块,一种称为关键数据块 (`critical chunk`),这是标准的数据块,另一种叫做辅助数据块 (`ancillary chunks`),这是可选的数据块。关键数据块定义了 4 个标准数据块,每个 PNG 文件都必须包含它们,PNG 读写软件也都必须要支持这些数据块。虽然 PNG 文件规范没有要求 PNG 编译器对可选数据块进行编码和译码,但规范提倡支持可选数据块。

文中提出的第一种加密思路—文件头加密,就是对 PNG 图片的文件头进行隐写,通过加入自己构建的信息块^[13-15] (`chunk`),达到加密效果。将加密信息封装成 PNG 结构的信息块,加入该 `chunk` 后对 PNG 文件对的数据块重新排序写入文件。

用算法 1 总结这个过程。

算法 1:PNG 文件头加密请求 `QueryEn`。

输入:PNG 格式文件,自定义加密信息;

输出:加入加密信息后的 PNG 文件。

- (1) 将待加密信息封装成固定格式的数据块;
- (2) 判断文件的可加密容量是否够用,不够则抛出异常;
- (3) 按字节读入图片,将图片划分为独立的数据块 (`PngChunk`);
- (4) 将现有的数据块 (含加密信息块),按照特定的顺序重新写入文件,并返回加密成功文件 `ΔData`。

第二种加密思路是像素点加工加密^[16-17],用 `java` 语言中的 `image` 类对 PNG 图片的像素点 `rgb` 值进行读写,将加密信息每 3 Byte 组成一个像素点,写入到图片中,实现嵌入有意义的信息,同时不影响 PNG 的使用,保护了文件的完整性,用对图片最小的破坏,换取最大的加密效果。

算法 2:像素点加密请求 `QueryEn`^[18-20]。

输入:PNG 格式文件,自定义加密信息;

输出:加入加密信息后的 PNG 文件。

- (1) 将待加密信息封装成固定格式的数据块;
- (2) 判断文件的可加密容量是否够用,不够则抛出异常;
- (3) 按像素读入图片,按特定的间隔取出对应的像素;
- (4) 将加密数据块按 3 Byte 每组进行划分,对每一组字节,前面添加 `0xff`,然后转换成 32 bit 的 `int`;
- (5) 将 `int` 值写入前面取出的像素值中,然后将像素重新写回文件,并返回加密成功文件 `ΔData`。

2.2 TXT 格式的加密实现

文本文件是基于字符编码的文件,常见的编码有 ASCII、UNICODE 等等。首先,读取文件上的二进制比特流,然后按照所选择的解码方式来解码,最后显示结果。一般来说,选取 ASCII 码作为解码形式。

根据文本文件的特点,构建了两种加密思路。第一种,在 TXT 文件的每行的行尾加密^[21-22],追加一定位数的不可见字符串 (如 `0x00` 和 `0x20`),该字符串由加密信息的二进制编码而来。第二种,在文本的末尾追加^[23-24] 一定行数的不可见字符串,该字符串由加密信息的二进制编码而来。

算法 3:行尾加密算法 `QueryEn`。

输入:TXT 格式文件及加密信息 `K`;

输出:加密后的 TXT 文件。

- (1) 按字节读入图片,并分割为独立的数据块;
- (2) 匹配数据块的块名,匹配成功则取出这个数据块;
- (3) 将封装好的加密信息转换成对应的 2 进制字

符串,并用特定字符(0x00 和 0x20)进行编码;

(4)按行读入文本文件,在行尾追加一定数量的加密字符后写回文件中。

算法 4:文本末尾加密算法 QueryEn。

输入:TXT 格式文件及加密信息 K;

输出:加密后的 TXT 文件。

(1)按字节读入图片,并分割为独立的数据块;

(2)匹配数据块的块名,匹配成功则取出这个数据块;

(3)将封装好的加密信息转换成对应的 2 进制字符串,并用特定字符(0x00 和 0x20)进行编码;

(4)在文本文件末尾追加一定行数的加密字符后写回文件。

采用对文本每行行尾加密的方法时,对文本的行数有一定的要求。目前定的标准是,每行行尾追加 5 Byte 长的二进制化的消息,用的是 0x00 和 0x20 编码,显示为空格,也就是说,行数要在 8 行以上。采用对文本最后追加消息加密的方法时,对每行的加密位数限制为 100 bits。上述信息长度标准都可以在程序中直接修改,不会对加密系统整体造成影响。

对 TXT 文件,同样构建了解密算法,方便对泄露文件追踪泄密源。

算法 5:解密 TXT 文件算法 QueryDe。

输入:加密过的 TXT 文件 ΔData;

输出:加密信息 K。

行尾解密:

(1)按行读入文本文件,从末尾开始,取出指定数目的字符;

(2)将取得的字符进行拼接,转换成二进制字符,再转换成特定格式的数据块;

(3)对数据块内容进行校验,每异常就取出加密信息。

文件末尾解密:

(1)直接读取文件末尾的加密信息,以空换行为分割符;

(2)将取得的加密信息进行拼接,转换成二进制字符,再转换成特定格式的数据块;

(3)对数据块内容进行校验,每异常就取出加密信息。

3 结果与分析

根据上述 INVIEN 系统的构造描述,对该方案成功地进行了实现。该方案主要包括初始化,输入,加密和解密四个过程。

3.1 PngOperation 模块性能与测试

在 PngOperation 模块中,采用了文件头加密和像素点加密两种实现方法。

在文件头加密方法和像素点加密方法中,按照等比数列分布,分别取了 1 M,2 M,4 M,8 M,16 M,32 M,64 M,128 M,256 M 数据量大小的 PNG 图片文件进行测试。

显然,像素点加密和文件头加密都是针对 PNG 格式的图片进行处理的,但在处理的速度和性能上,还是存在一定差距的。故在多次计算,取其平均值后,可以看到该方案针对批量文件的去尾平均时间如图 3 所示。

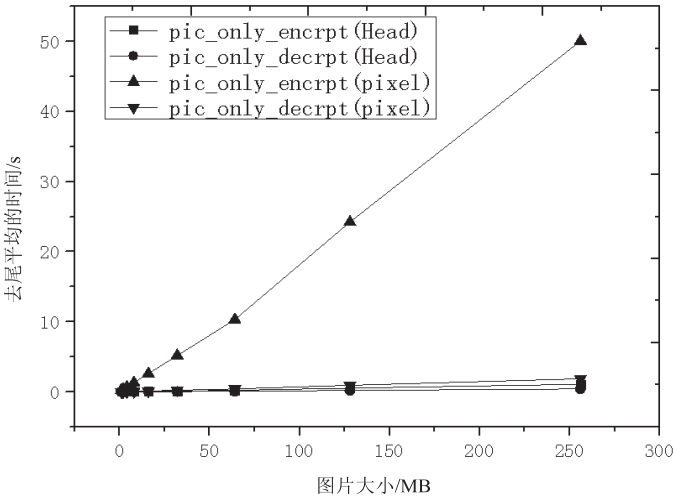


图 3 PngOperation 模块两种不同隐写方式的加密解密时间曲线

图 3 中,pic_only_encrypt(Pixel) 曲线表示的是像素点加密方法,pic_only_encrypt(Head) 曲线表示的是文件头加密方法,pic_only_decrypt(Head) 曲线表示的是文件头解密,pic_only_decrypt(Pixel) 曲线表示的是

像素点解密。显而易见,pic_only_encrypt(Pixel) 与其他三条曲线差异最大,且接近参考的指数曲线。换句话说,像素点加密方法的时间是随着图片文件的增加呈指数增长的,而文件头加密和两者所对应的解密方

法所需的时间都是缓慢增长的。

3.2 TXTOperation 模块性能与测试

在 TXTOperation 模块中,采用了行尾加密和文本末尾追加加密两种实现方法。

在行尾加密方法和文本末尾追加加密方法中,考虑到 TXT 文本文件小的可以只有几 k,而文件大的有

几百 M。因此,对小文件和大文件分别讨论,并按照等比数列分布,分别取了由 1 个,2 个,4 个,8 个,16 个,32 个,64 个,128 个加密的气象文本数据样本构成的文件进行测试。在多次计算取其平均值后,可以看到该方案针对批量文件的去尾平均时间如图 4 和图 5 所示。

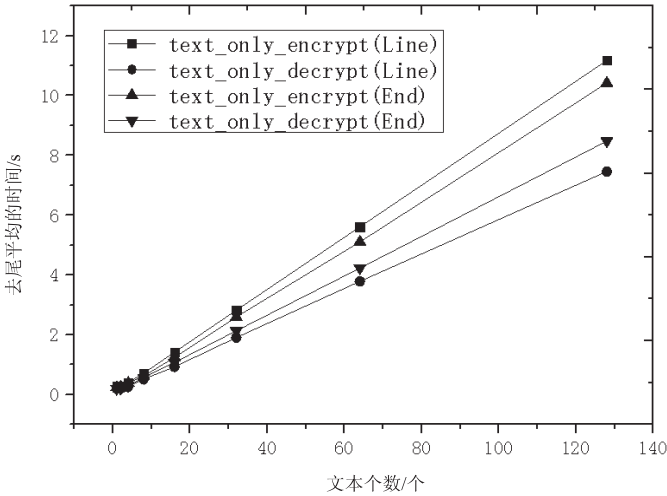


图 4 TXTOperation 模块在小文本文件下的两种不同隐写方式的加密解密时间曲线

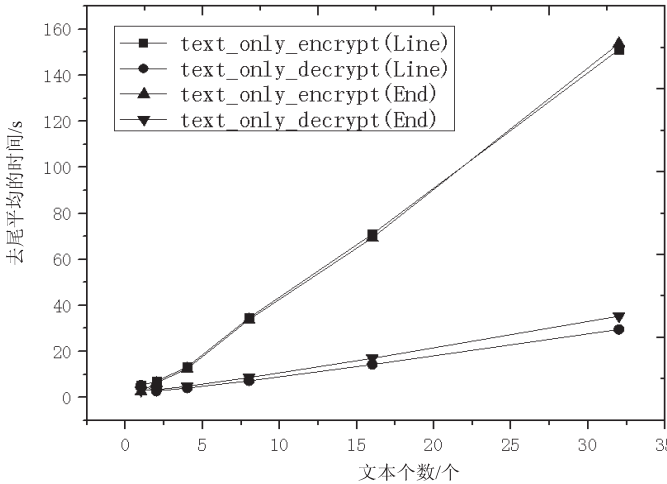


图 5 TXTOperation 模块在大文本文件下的两种不同隐写方式的加密解密时间曲线

在图 4 和图 5 中, text_only_encrypt(End) 曲线表示的是文本末尾追加加密方法, text_only_encrypt(Line) 曲线表示的是行尾加密方法, text_only_decrypt(Line) 曲线表示的是行尾解密, text_only_decrypt(End) 曲线表示的是文本末尾追加解密。

由图 4 可知,针对小文本文件,无论是行尾加密方法,还是文本末尾追加加密方法,其加密时间都是随着文本数量的增加呈指数增长的,其对应的解密时间也同样呈指数增长。但总体说来,行尾加密的时间大于文本末尾追加加密的时间,那么相反的,行尾解密的时间小于文本末尾追加解密的时间。

由图 5 可知,针对大文本文件,曲线比较像 PngOperation 模块中,图 3 对应的两种方法加解密的

时间走势。具体分析, text_only_encrypt(End) 与其他三条曲线差异最大,且接近参考的指数曲线。换句话说,文本末尾追加加密方法的时间是随着文本文件个数的增加呈指数增长的,而行尾加密和两者所对应的解密方法所需的时间都是缓慢增长的。

3.3 性能测试结果与分析

3.3.1 PngOperation 模块

可以看到,两种方案的解密速度都是很快的,但是加密时间却差异很大。随着图片文件数据量的不断增大,像素点加密的时间开始急剧增加,而文件头加密所需之间仍只是接近线性增长。

总的说来,针对 PNG 的图片文件而言,文件头加密是相对较快的,尤其是对于图片文件较大的时候。

但是,也不得不考虑整个隐写算法的安全性,从这点出发,文件头加密的鲁棒性能就较差了,加密的用户信息很容易被去除。而像素点加密的算法,虽然可能会造成对原始数据的轻微改动,但是隐写的用户信息很难察觉,并且也并不容易去除。

3.3.2 TXTOperation 模块

可以看到,两种方案的解密速度相较于对应的加密速度更快,并且两种方案的加密时间也相差无几。

在 PngOperation 模块性能与测试中,已经看到了像素点加密和文件头加密在一定情况下,是有显著区别的。那么针对 TXT 文本文件的两种方法,一个是在行尾加密,一个是在文本末尾追加加密,为什么加密的时间也相差无几,其实就是因为这两种方法本质加密的原理是一样的,只是隐写嵌入加密的用户信息时,嵌入位置不一样而已。

所以,总的说来,针对 TXT 的文本文件而言,两种方法的效率是差不多的。文中方法是将加密的用户信息转成二进制使用 winhex 工具查看,并追加使用 0x00 和 0x20 构成的一定位数的不可见字符串构成,因此行尾追加的方法会比文本末尾追加的方法的鲁棒性和安全性更好,所以优先考虑行尾追加方法。但由于文本文件没有固定的格式,若原始的文本文件内容很少,行数也只有几行,那么行尾追加的方法是不能将所有经过加密的用户信息成功嵌入的。因此,这时直接在文本末尾追加完整的经过加密的用户信息,就不用受行数约束。

4 结束语

该课题主要基于隐写技术进行研究,通过搭建气象数据源版权保护系统平台,可以对气象数据添加隐写用户信息。当数据在开放的网络环境下共享后,如发现违规操作和数据泄露情况时,通过该系统可以对气象数据对外共享后的数据鉴别提供支撑依据,从而实现对气象数据的版权保护。

参考文献:

- [1] 黄方军,黄继武. 基于图像校准的通用型 JPEG 隐写分析[J]. 中国科学,2009,39(4):383-390.
- [2] 王 鹏. 基于 JPEG 图像的信息隐藏系统[D]. 郑州:解放军信息工程大学,2012.
- [3] 魏鹤君,吕笑倩,郑彩平. 基于 RSA 的数字签名技术研究[J]. 网络安全技术与应用,2007(8):43-44.
- [4] 卡曾贝塞,佩蒂科勒斯. 信息隐藏技术:隐写技术与数字水印[M]. 北京:人民邮电出版社,2001.
- [5] 李 星. JPEG 图像及解压图像中的隐写分析技术研究[D]. 郑州:解放军信息工程大学,2012.
- [6] 朱晓冬,苑森森,刘 静,等. 空域数字水印算法[J]. 吉林大学学报:工学版,2003,33(2):56-59.
- [7] 孙圣和,陆哲明. 数字水印处理技术[J]. 电子学报,2000,28(8):85-90.
- [8] 吕晓敏. 嵌套循环冗余码(CRC)的优化与检验[D]. 杭州:浙江大学,2012.
- [9] 洪丹丹,罗军峰,冯兴利,等. 基于 RSA 与 MD5 签名的实名制微门户设计[J]. 微电子学与计算机,2016,33(9):36-41.
- [10] 李夏梦,潘广贞. 基于消息摘要算法第五版和 IDEA 的混合加密算法[J]. 科学技术与工程,2017,17(9):233-238.
- [11] 罗向阳. 数字图像隐写检测关键问题研究[D]. 郑州:解放军信息工程大学,2010.
- [12] 王朔中,张新鹏,张卫明. 以数字图像为载体的隐写分析研究进展[J]. 计算机学报,2009,32(7):1247-1263.
- [13] 毛家发,林家骏,戴 蒙. 基于图像攻击的隐藏信息盲检测技术[J]. 计算机学报,2009,32(2):318-327.
- [14] 殷赵霞,汤 进,刘燕君,等. 基于像素对匹配的高载荷隐写算法[J]. 系统工程理论与实践,2013,33(11):2972-2979.
- [15] 孙文颀,刘婷婷,张新鹏,等. 彩色图像通用隐写分析的多类统计特征[J]. 中国图象图形学报,2008,13(10):1914-1917.
- [16] 袁文翀. 数字隐写图像的特征提取及信息检测方法[D]. 南昌:南昌大学,2012.
- [17] 陈 丹,王育民. 一种针对加性空域掩密算法的通用掩密分析技术[J]. 东南大学学报:自然科学版,2007,37:48-52.
- [18] 熊 钢. 基于图像区域统计特征的隐写分析技术研究[D]. 郑州:解放军信息工程大学,2012.
- [19] 万宝吉. 基于融合的图像隐写分析技术研究[D]. 郑州:解放军信息工程大学,2013.
- [20] 李开达,张 涛,李 星. 基于模糊积分多分类器融合的 JPEG 图像隐写算法识别[J]. 信息工程大学学报,2012,13(2):200-204.
- [21] 黄 伟,李 洪,高 健. 基于文本隐写技术的档案源头追踪模型研究[C]//全国档案工作者年会. 福建,厦门:中国档案学会,2014:3-5.
- [22] 金 烨,眭新光. 新的文本隐写分析方法[J]. 计算机工程,2008,34(21):159-160.
- [23] 向凌云,王鑫辉. 分级安全的文本隐写方法[J]. 计算机应用,2015,35(3):717-721.
- [24] 武莉莉. 一种基于图像处理的防篡改脆弱文本数字水印算法[J]. 科学技术与工程,2013,13(25):7563-7567.