安全的密文域图像隐写术

陈嘉勇*①② 王 超① 张卫明① 祝跃飞①
①(解放军信息工程大学信息工程学院 郑州 450002)
②(中国科学院信息安全国家重点实验室 北京 100049)

摘 要：基于同态加密和双层隐写编码，该文提出一种安全的密文域图像隐写术，其可以达到传统明文隐写术的容量，并且在密文域和明文域均能有效抵抗隐写检测分析。首先结合自适应隐写术和湿纸编码技术，提出一种明文域双层隐写算法；其次，修正一种全同态加密算法，对载密图像进行加密；最后，在密文域上提取嵌入的信息。理论分析和实验结果表明：在加密/隐写密钥同时泄露、加密密钥泄露和密钥未泄露条件下，算法均具有较高的安全性。

关键词：隐写术；自适应编码；湿纸编码；同态加密
中图分类号：TP309.7 **文献标识码**：A **文章编号**：1009-5896(2012)07-1721-06
DOI: 10.3724/SP.J.1146.2011.01240

A Secure Image Steganographic Method in Encrypted Domain

Chen Jia-yong①② Wang Chao① Zhang Wei-ming① Zhu Yue-fei①
①(*Information Science and Technology Institute*,
PLA Information Science and Technology University, *Zhengzhou* 450002, *China*)
②(*State Key Laboratory of Information Security*, *Graduate University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)

Abstract: A new secure image steganographic algorithm for encrypted domain steganography is proposed based on fully homomorphic encryption and double-layer embedding. The proposed method can reach the capacity of classical steganography while resist to statistical steganalysis both in plaintext domain and encrypted domain. At first, a novel double-layered embedding algorithm based on adaptive steganography and Wet Paper Codes(WPCs) is constructed for plaintext embedding. Then an fully homomorphic encryption algorithm is modified to encrypt the stego image. Finally, the embedding message is extracted in encrypted domain. The performance analysis and experimental results show that the steganographic security of the proposed method preserves well performance under several attacking conditions, including both steganographic/encrypted keys leakage, encrypted keys leakage and no keys leakage.

Key words: Steganography; Adaptive coding; Wet Paper Coding (WPC); Homomorphic encryption

1 引言

云计算是当前信息技术领域的研究热点之一。随着云计算的不断普及，用户对数据安全性与隐私性的需求导致安全问题已成为制约云计算发展的重要因素[1]。加密算法作为云安全的重要基石，被广泛用于保护用户的数据安全性和隐私性。然而，加密算法的应用也对云服务提出了新的挑战：数据变成密文后丧失了原有特性，导致目前大部分数据处理方法失效。密文域的信号处理问题是云安全的关键问题之一，而同态加密技术是实现密文域信号处理的重要基础，其允许用户对密文数据直接进行运算而不影响其保密性。例如，2009年Gentry[2]提出的基于多项式环上“理想格”的全同态算法与2010年Dijk等人[3]提出的针对整数加密的全同态加密算法是两个典型算法。

信息隐藏是将消息嵌入到多媒体数据中的一种技术，可用于隐蔽传输或版权保护等目的。随着数据隐私保护需求的日益强烈，如何实现密文域信息隐藏也成为研究热点之一。根据数据加密和数据嵌入的结合方式，密文域信息隐藏主要可分为局部加密和全局加密两种方式。其中，局部加密方式将载体数据分为两部分，其中一部分用于数据加密，剩余部分载体用于负载嵌入信息。例如：文献[4]将变换域上的载体数据分为高低位平面，分别进行加密和嵌入水印信息。文献[5]为每个消息拥有者生成不同的指纹信息，消息的拥有者将载体离散余弦变换

2011-11-28 收到，2012-03-22 改回
国家自然科学基金(60803155, 61170234, 60970141, 60902102)和郑州市科技创新团队项目(10CXTD150)资助课题
*通信作者：陈嘉勇 cjy1003@sina.com

(DCT)系数的符号进行加密，而每个载体使用者使用不同的密钥解密一部分系数。局部加密模式的主要缺点是容易造成载体信息的泄露，而全局加密方式则能更好的保证数据的保密性。文献[6–8]分别基于Okamoto-Uchiyama, Paillier和Chameleon等加密体制的同态性质，在载体信号中嵌入额外的数据信息。这类方法目前主要用于在买方-卖方水印协议中实现水印的稳健性、不可见性、公平性及匿名性等问题，其主要缺点在于密文的码率扩张大且计算复杂度较高。2011年，Zhang[7]提出针对流密码加密算法的可逆图像隐写算法，该方法具有很高的计算效率和保密安全性。然而其隐藏容量较低，且在攻击者获取解密密钥条件下，利用最低有效位(LSB)隐写检测算法即可检测隐蔽信息的存在性。

上述密文域信息隐藏的研究主要集中在密文域数字水印和密文域可逆隐藏，其目的是用于版权保护、数据完整性认证或数据标注。但是如何以密文为载体进行隐蔽通信——即密文域隐写方面的研究还很少。一方面，传统的阈下信道技术[8]以网络通信协议为掩护传输信息，具有可证明安全性，但是其信息传输率很低，比如基于密钥交换协议的阈下信道，每次交互只传输几比特秘密信息。另一方面，传统的以明文多媒体数据载体的隐写术可以提供非常高的嵌入容量，但是难以做到安全。

云环境下，为了使用云服务不泄露数据隐私，用户需要对数据进行加密，并希望云能在密文域完成数据分析，比如密文域图像检索或聚类，为此需要采用同态加密、保序加密等特殊的加密方法。云环境下密文域信号处理的推广事实上为在加密多媒体数据上进行隐蔽通信提供了良好的平台。这种隐蔽通信以某些密文域信号处理协议为掩护，故可以看成是阈下信道的推广，本文称这种隐蔽通信方式为密文域隐写。本文将对密文域隐写方式进行分类，基于同态加密和双层隐写编码给出一种密文域隐写方案，分析表明这种隐写术可以达到传统明文隐写术的容量，并且在密文域和明文域均能有效抵抗隐写检测分析。

2 密文域隐写算法

2.1 记号

下面用用斜体字母表示变量，大写黑体字母表示矩阵，小写黑体字母表示向量。令 Σ 为有限字符集，则 Σ^* 和 Σ^∞ 分别表示有限序列和无限序列。记序列 s 的长度为 $|s|$。序列 s_1 和 s_2 的连接记为 $s_1 \| s_2$。对整数 z，分别用 $\lceil z \rceil$，$\lfloor z \rfloor$ 和 $\langle z \rangle$ 表示 z 的向上取整、向下取整和四舍五入取整。

2.2 密文域隐写算法

本文主要研究明文域嵌入-密文域提取(embedding in Plaintext-extracting in Ciphertext)，即 P2C 模式环境下如何进行安全的隐蔽通信。总体思路如下：首先，利用双层嵌入方法，在明文域载体中嵌入数据；其次，修正一种全同态加密算法，对载密图像进行加密；最后，在密文域上提取嵌入的信息。

(1)明文域双层嵌入　借鉴文献[9]的双层嵌入编码思想，本文提出一种适用于明文域隐写的双层嵌入方法。首先，在载体的 LSB 层采用自适应编码嵌入消息，并需要修改的像素标记为“干”，无需修改的像素点标记为“湿”。其次，在次 LSB 层用湿纸码嵌入消息，两层需要的修改用+1 或–1 的方式同时完成。

不失一般性，假设载体 C 为 $M \times N$ 的 8 bit 灰度图像，记为 $C = \{c_{ij}\}$。其中，$1 \le i \le N$，$1 \le j \le M$，$c_{ij} \in [0,255]$，取 $n = M \times N$。首先，根据纹理复杂度定义每个像素点的失真度量，记像素点 c_{ij} 的纹理复杂度 $\overline{\rho}(c_{i,j})$ 为

$$\overline{\rho}(c_{i,j}) = \big((c_{i,j} - c_{i,j-1})^2 + (c_{i,j} - c_{i,j+1})^2 + (c_{i,j} - c_{i-1,j})^2 + (c_{i,j} - c_{i+1,j})^2\big)\big/4 \quad (1)$$

定义载体 C 的失真度量矩阵 $\boldsymbol{\rho}(C) = \{\rho(c_{i,j})\} = \{1/(\overline{\rho}(c_{i,j}) + 1)\}$，即纹理越复杂则修改代价越小。记嵌入率为 α，记载体纹理复杂度最小的 $(1-2\alpha)n$ 个像素的集合为 $C_0 = \{c_{i,j}\}$，取 $\rho(c_{i,j} \mid c_{i,j} \in C_0) = \infty$。以 $\boldsymbol{\rho}(C)$ 为边信息，利用文献[10]提出的 STC 编码方法在 LSB 层实现自适应隐写。在得到 LSB 层的预期修改位置后，并不直接修改载体像素值，而是把需要修改的像素标记为“干”，无需修改的像素点标记为“湿”，从而次 LSB 层构造了一条新的湿纸隐写信道。采用文献[11]中的 LT 方法，在次 LSB 层的湿纸信道上完成消息的嵌入。接收方对载密图像的 LSB 序列和次 LSB 分别执行 SCT 的解码算法和湿纸码的解码算法即可提取消息。上述嵌入方法既避开了容易受统计检测攻击的载体图像敏感区域(即 $\rho(c_{i,j})$ 取值较大的区域)，同时具有较大嵌入容量。

(2)全同态加密　下面基于文献[3]的思想进行修正，构造一种新的全同态加密算法，新算法既保持同态性，能够完成密文域上的信号处理，还可用于隐蔽通信。取 $k = 8$，$\tau = 2$，算法安全参数为 λ(通常可取几十到几百比特)，构造加密算法如下：

(a)密钥生成算法 KeyGen(λ)：用于生成加密密

钥 p。其中，$p \leftarrow [2^{\lambda-1}, 2^{\lambda})$，且 p 为奇数。

(b) 加密算法 $\mathrm{Enc}(p, X)$：记明文 $X = \{x_i \mid x_i \in [0, 2^k]\}$，密文 $Y = \{y_i\}$。取 $w_i = pq_i + 2^k r_i (i = 0,1,\cdots,n)$，其中，$q_i$，$r_i$ 为随机数，$r_i \approx 2^{\sqrt{\lambda}}$，$q_i \approx 2^{\lambda^3}$。公开 q_0 和 r_0，令 $q_i = 2^\tau \lfloor q_i / 2^\tau \rfloor (i = 1,2,\cdots,n)$，则

$$\begin{aligned}\mathrm{Enc}(p, x_i) &= (w_i + x_i) \bmod w_0 \\ &= \left(pq_i + 2^k r_i + x_i\right) \bmod \left(pq_0 + 2^k r_0\right) = y_i\end{aligned}$$

(c) 解密算法 $\mathrm{Dec}(p, Y)$：输出 $x_i = (y_i \bmod p) \cdot \bmod 2^k$。

对模加运算有：$\mathrm{Enc}(x_i) + \mathrm{Enc}(x_j) = (p(q_i + q_j) + 2^k(r_i + r_j) + (x_i + x_j)) \bmod (pq_0 + 2^k r_0)$。此时，$\mathrm{Dec}(p, \mathrm{Enc}(x_i) + \mathrm{Enc}(x_j)) = \mathrm{Dec}(p, \mathrm{Enc}(p, x_i + x_j))$。

对模乘运算有：$\mathrm{Enc}(x_i) \cdot \mathrm{Enc}(x_j) = (p \cdot f_p(q_i, q_j, r_i, r_j, x_i, x_j) + 2^k \cdot f_r(r_i, r_j, x_i, x_j) + (x_i \cdot x_j)) \bmod (pq_0 + 2^k r_0)$。其中，$f_p(q_i, q_j, r_i, r_j, x_i, x_j), f_r(r_i, r_j, x_i, x_j) \in Z^+$。

此时，$\mathrm{Dec}(p, \mathrm{Enc}(x_i) \cdot \mathrm{Enc}(x_j)) = \mathrm{Dec}(p, \mathrm{Enc}(p, x_i \cdot x_j))$。

综上，上述加密算法一方面可用于密文域信号处理，另一方面也为隐蔽通信提供了通信信道。

(3)密文域隐写　下面考察消息发送方在明文域嵌入消息后，接收方在密文域提取消息的正确性。

记 $\mathrm{LSB}(x) = x\&0\mathrm{x}00\cdots01$，$2\mathrm{LSB}(x) = x\&0\mathrm{x}00\cdots02$，由于 $q_i = 2^\tau \lfloor q_i / 2^\tau \rfloor$，故

$$\begin{aligned}\mathrm{Enc}(p, x_i) = &\left(2^\tau p \lfloor q_i / 2^\tau \rfloor + 2^k r_i + x_i\right) \\ &- t_0\left(2^\tau p \lfloor q_0 / 2^\tau \rfloor + 2^k r_0\right) \qquad (2)\end{aligned}$$

因此，对密文 y_i，有

$$\mathrm{LSB}(y_i) = \mathrm{LSB}(\mathrm{Enc}(p, x_i)) = \mathrm{LSB}(x_i) \qquad (3)$$

$$2\mathrm{LSB}(y_i) = 2\mathrm{LSB}(\mathrm{Enc}(p, x_i)) = 2\mathrm{LSB}(x_i) \qquad (4)$$

其中，式(3)，式(4)表明，接收方无需解密，即可从密文数据的 LSB 层和次 LSB 层正确提取秘密信息。

下面给出具体的 P2C 模式下的密文域隐写算法如表 1 所示。

表 1 P2C 模式密文域图像隐写算法

初始状态：发送方拥有灰度图像 $C = \{c_{i,j}\}$，消息 $M = \{m_i\}$，隐写密钥 p'，加密密钥 p；接收方拥有隐写密钥 p'。

(1)嵌入算法：

步骤 1　预处理。计算载体图像 C 的失真矩阵 $\rho(C) = \{\rho(c_{i,j})\}$。用隐写密钥 p' 对 C 进行随机置乱，记置乱后的载体为 $\mathrm{Per}(C)$。

步骤 2　消息嵌入。用明文域嵌入算法在 $\mathrm{Per}(C)$ 中嵌入消息 M 得载密图像 S。

步骤 3　数据加密。用同态加密算法对载密图像 S 进行加密，得到加密后的载密图像 Y，将 Y 发给服务器。

(2)提取算法：

步骤 1　预处理。接收方从服务器下载加密图像 Y。用隐写密钥 p' 对 Y 进行逆置乱，记逆置乱后的载体为 $\mathrm{iPer}(Y)$。

步骤 2　消息提取。采用密文域提取算法提取 $\mathrm{iPer}(Y)$ 的 LSB 序列和次 LSB 序列。对 LSB 序列采用 STC 解码算法提取秘密信息，对次 LSB 序列采用 LT 解码算法提取秘密信息。

3　安全性分析

假设同态加密算法 Φ 和隐写算法 Φ' 对所有人是公开的，通信双方 Alice 和 Bob 在同态加密的隐蔽信道中采用密文域隐写系统 $\{\Phi', \Phi\}$ 进行隐蔽通信，Wendy 是信道上的攻击者。下面证明：在 Wendy 无密钥信息、Wendy 获得加密密钥、Wendy 获得加密密钥与隐写密钥 3 种条件下，本文算法都具有很高的安全性。这里的安全性在此特指隐蔽安全性，即系统抗隐写检测攻击的能力。

3.1 安全性定义

下面讨论密文域隐写系统 $\{\Phi', \Phi\}$ 的安全性。首先定义基本喻示。

(1)加密/解密喻示。加密喻示 O_{enc} 和解密喻示 O_{dec} 都是由同态加密的密钥 p 控制的。其中，O_{enc} 的输入是明文，输出是对应的密文；O_{dec} 的输入是密文，输出是对应的明文或失败。

(2)置乱/逆置乱喻示。置乱喻示 O_{per} 和逆置乱喻示 O_{iper} 都是由隐写密钥 p' 控制的。其中，O_{per} 的输入是明文域载体，输出是置乱后的明文域载体；O_{iper} 的输入是置乱后的明文域载体，输出是明文域载体或失败。

(3)隐写检测喻示。隐写检测喻示 $O_{\mathrm{det\_}\Phi'}$ 的输入是 $X \in \{\Omega, \Omega^*\}$，其中 Ω 表示载体对象 Ω，Ω^* 表示载密对象，输出 $\mathrm{Det}(X) \in \{0,1\}$，其中 0 表示 X 未载密，1 表示 X 载密。

其次，定义系统攻击者。对 $\{\Phi', \Phi\}$ 定义 3 种攻击者：

(1)IND-CSA　(INDistinguish-Cipher-Stego-only Attack)攻击者。IND-CSA 攻击者拥有加密后的载密图像 S，其允许访问加密/解密喻示，置乱/逆置乱喻示和隐写检测喻示，并据此判断 S 是否载密。

(2)IND-KEA　(INDistinguish-Known-Encrypt-key Attack)攻击者。IND-KEA 攻击者拥有加密后的图像 S 和加密密钥 p，其允许访问加密/解密喻示，置乱/逆置乱喻示和隐写检测喻示，并据此判断 S 是否载密。

(3)IND-KBA　(INDistinguish-Known-Both-keys Attack)攻击者。IND-KBA 攻击者拥有加密后的图像 S、加密密钥 p 和隐写密钥 p'，其允许访问加密/解密喻示，置乱/逆置乱喻示和隐写检测喻示，并据此判断 S 是否载密。

此外，针对 Φ' 定义 1 种攻击者：

IND-DA (INDistinguish-Detecting Attack) 攻击者。IND-DA 攻击者可对明文域图像 S 判断其是否载密。

定义 1 Φ' 的 IND-DA 检测攻击者 A 定义为一个概率多项式时间算法 A，A 可访问检测喻示 $\{O_{\det\_\Phi'}\}$。

对未知对象 $X\in\{\Omega,\Omega^*\}$，隐写检测者的任务是回答 X 是否载密。借鉴 Barbier 等人[12]的隐写安全性模型，采用攻击算法的检测性能(包括漏检率 α 和虚警率 β)刻画隐写检测者的能力。其中，漏检率 $\alpha=p(\Omega\,|\,\Omega^*)$，虚警率 $\beta=p(\Omega^*\,|\,\Omega)$，且 $0\le\alpha,\beta\le1$。

定义 2 $\{\Phi',\Phi\}$ 的 IND-CSA 攻击者 A 是一对概率多项式时间算法 (A_1,A_2)，A_i 可访问喻示 $O_i(i=1,2)$，其中 O_1 包含 $\{O_{\text{enc}},O_{\text{dec}}\}$，$O_2$ 包含 $\{O_{\text{enc}},O_{\text{dec}},O_{\det\_\Phi'}\}$。

定义 3 $\{\Phi',\Phi\}$ 的 IND-KEA 攻击者 A 是一对概率多项式时间算法 (A_1,A_2)，A_i 可访问喻示 $O_i(i=1,2)$，其中 O_1 包含 $\{O_{\text{enc}},O_{\text{dec}},O_{\text{per}},O_{\text{iper}}\}$，$O_2$ 包含 $\{O_{\text{enc}},O_{\text{dec}},O_{\det\_\Phi'}\}$。

定义 4 $\{\Phi',\Phi\}$ 的 IND-KBA 攻击者 A 是一对概率多项式时间算法 (A_1,A_2)，A_i 可访问喻示 $O_i(i=1,2)$，其中 O_1 包含 $\{O_{\text{enc}},O_{\text{dec}},O_{\text{per}},O_{\text{iper}}\}$，$O_2$ 包含 $\{O_{\text{enc}},O_{\text{dec}},O_{\text{per}},O_{\text{iper}},O_{\det\_\Phi'}\}$。

$\text{Exp}_{\Sigma}^{\text{IND-ATK}}(A,k)$ 指攻击者 A 在 ATK 条件下对系统 Σ 进行区分攻击的实验。对本文而言，$\text{ATK}\in\{\text{CSA,KEA,KBA}\}$，系统 $\Sigma\in\{\Phi',\Phi,\{\Phi',\Phi\}\}$，安全参数 $k\in\{\lambda,\lambda',(\lambda,\lambda')\}$。

定义 5[13] 在安全参数为 k 条件下，记 IND-ATK 攻击者对系统 Σ 的攻击优势为 $\text{Adv}_{\Sigma}^{\text{IND-ATK}}(A,k)$：

$$\text{Adv}_{\Sigma}^{\text{IND-ATK}}(A,k)=2\left|\Pr\left(\text{Exp}_{\Sigma}^{\text{IND-ATK}}(A,k)=1\right)-\frac{1}{2}\right|$$

其中 $\text{Adv}_{\Sigma}^{\text{IND-ATK}}(A,k)$ 取值大小与攻击者的能力强弱成正比。

定义 6 称系统 Σ 在 IND-ATK 条件下是 ε-安全的，当且仅当 $\text{Adv}_{\Sigma}^{\text{IND-ATK}}(A,k)\le\varepsilon$。特别地，当 $\varepsilon=0$ 时，Σ 是绝对安全的。

3.2 明文域抗检测分析

首先，考虑本文提出的算法在明文域的抗检测性能。其中，导致 IND-KBA 攻击的可能情形是：服务器在紧急状况下强制要求通信双方对通信数据进行解密和逆置乱，并在明文域检测图像数据是否存在秘密信息。显然，此时密文域隐写系统 $\{\Phi',\Phi\}$ 的安全性仅取决于明文域隐写算法的安全性。

采用 UCID 图像库[14]中规格为 384×512 的 1000 幅图像，用 LSBM(LSB Matching)算法、Luo 等人[15]提出的 EA(Edge Adaptive)算法、Lu 等人[16]提出的 NRE(Noisy Region Embedding)算法和本文算法分别以嵌入率 α 嵌入消息。然后，对上述 4 种方法分别采用 Cancelli 等人[17]提出的 ALE 检测算法和 Cai 等人[18]提出的 RH 检测算法进行检测。在嵌入率 α 分别为 0.5 和 0.25 时，得到 ROC 曲线如图 1 所示。其中，检测的性能用 ROC 曲线进行评价，横坐标为虚警率，纵坐标为载密对象的正确检测率。

由实验结果可以看出，LSBM 的抗检测能力最差，其次是 EA 算法。嵌入率为 0.5 时，本文算法的抗检测性能是最好的，嵌入率为 0.25 时，本文算法和 NRE 算法性能基本一致。ALE 检测算法对 EA, NRE 和本文算法的检测效果较差，在嵌入率 0.25 时，EA 和 NRE 隐写算法性能基本相当，本文算法性能稍优于前两者，但较 LSBM 算法的抗检测性能好。

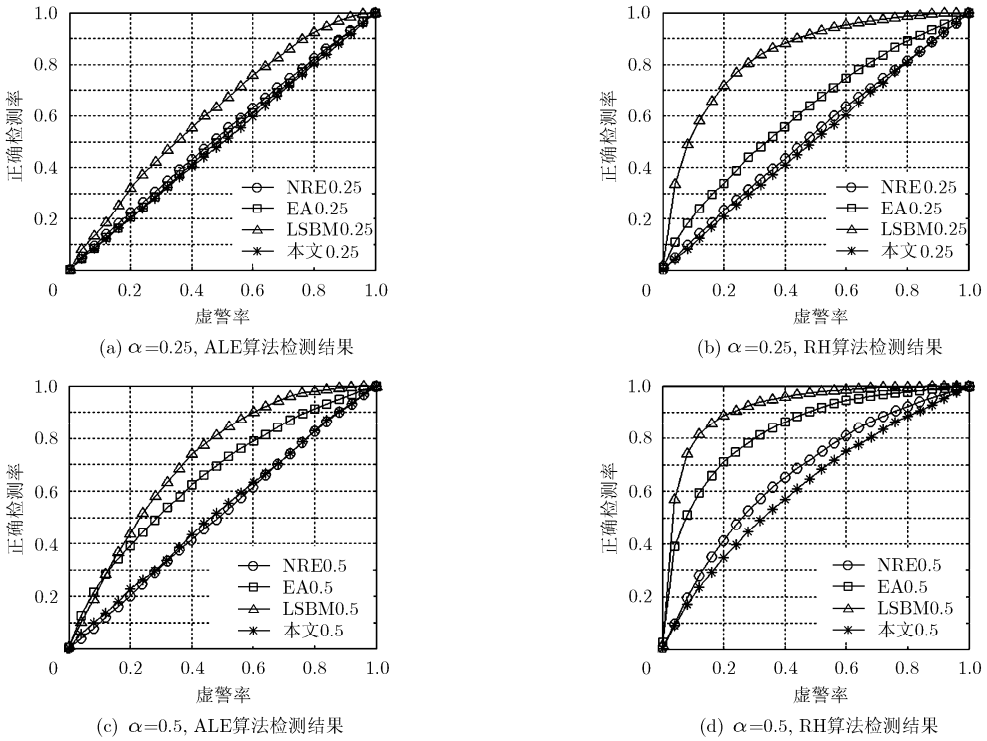
综上，在 IND-KBA 条件下，密文域隐写系统的安全性取决于隐写算法的安全性。此时，若 Φ' 是 ε'-安全的，则 $\{\Phi',\Phi\}$ 也是 ε'-安全的。

3.3 密文域安全性分析

下面考虑密文域隐写系统在密文域的抗检测性能，这里分别考虑存在 IND-KEA 攻击者和 IND-CSA 攻击者情形下的系统安全性。

(1) IND-KEA 攻击情形 首先，讨论存在 IND-KEA 攻击者的情形。其中，导致 IND-KEA 攻击的可能情况是：服务器通过中间人攻击等手段，获得加密密钥，其试图在没有隐写密钥情况下，检测图像数据是否含有秘密信息。

在实验 $\text{Exp}_{\{\Phi,\Phi'\}}^{\text{IND-KEA}}(A,\lambda')$ 的第 1 阶段，挑战者利用加密参数 λ 生成加密密钥 p，用隐写参数 λ' 生成隐写密钥 p'；攻击者用加密参数 λ 生成加密密钥 p，并可任意多次访问喻示 O_1，从消息集合 $\boldsymbol{M}$ 中任意选取待嵌信息 M。在第 2 阶段，挑战者从载体集合 $\boldsymbol{C}$ 中任意选取载体 C，并随机选择 1 bit b，若 $b=1$，则挑战者在 C 中嵌入消息 M 得到 S，利用 p' 对 S 进行置乱得到 $\text{Per}(S)$，加密 $\text{Per}(S)$ 得到 C^*，并将 C^* 返回给攻击者；若 $b=0$，利用 p' 对 C 进行置乱得到 $\text{Per}(C)$，利用加密算法将 $\text{Per}(C)$ 加密得到 C^*，将 C^* 返回给攻击者。攻击者可利用 p 对 C^* 进行解密，并结合 M 和 C^*，任意多次访问喻示 O_2，进而判断 C^* 是否载密，并将判断结果 b' 返回给挑战者。其中，若判断 C^* 载密，$b'=1$；否则，返回 $b'=0$。若攻击者判断正确，则 $\text{Exp}_{\{\Phi,\Phi'\}}^{\text{IND-KEA}}(A,\lambda')$ 取值为 1，否则取值为 0。



(a) α=0.25, ALE算法检测结果　(b) α=0.25, RH算法检测结果

(c) α=0.5, ALE算法检测结果　(d) α=0.5, RH算法检测结果

图 1 IND-KBA 攻击隐写检测结果

记 C^* 的低 2 层位平面数据为 $L(C^*)$，除低 2 层位平面外的数据为 $H(C^*)$。其中，$H(C^*)$ 部分的安全性仅取决于同态加密算法的安全性。记攻击者正确解密 C^* 的概率为 Pr(Dec_Success)。记 Φ' 是 ε'-安全的，则对 IND-KEA 攻击者而言，有 $\mathrm{Adv}^{\text{IND-KCA}}_{\{\Phi,\Phi'\}}(A,\lambda') = 2\left|\Pr\left(\mathrm{Exp}^{\text{IND-DA}}_{\Phi'}(A,\lambda')=1\right)-\frac{1}{2}\right| \le \varepsilon'$。

由于 IND-KEA 攻击者通过解密可得到 $H(C^*)$ 置乱后的数据，故可获得关于 $H(C^*)$ 的一阶统计信息。本文提出的加密算法本质上是局部加密算法，并未对 $L(C^*)$ 进行加密，故 $L(C^*)$ 抗检测攻击的能力取决于载体数据在低位平面的统计特性及置乱算法的安全性。根据文献[19]研究结果，载体的低 2 层位平面的统计特性类似随机噪声。考虑 IND-KEA 攻击者获得的是经过置乱后的载密图像的低 2 层位平面数据，其包含原始载体的部分低二层位平面信息和经由隐写编码嵌入的信息。由于前述明文域隐写算法采用隐写编码技术，根据湿纸编码和 STC 编码的性质，即使嵌入的明文信息具有明显的统计特性(如全 0 序列)，隐写编码后生成的密文序列依然呈现良好的伪随机统计特性，且在总体上保持 01 均衡。通过选用性能优良的置乱算法，发送方可将 01 均衡的载密位平面数据转化为伪随机序列。

根据明文域抗检测分析结果可知，在 IND-KBA 条件下，攻击者正确判断出载体是否载密的可能性接近于随机猜测。若 Φ' 是 ε'-安全的，则在 IND-KEA 条件下，由于攻击者只能获得关于载体高层位平面的部分统计特性，故此时 $\{\Phi',\Phi\}$ 至少是 ε'-安全的。因此，$\{\Phi',\Phi\}$ 在 IND-KEA 条件下具有相较于 IND-KBA 条件下更高的安全性。

(2)IND-CSA 攻击情形　最后，讨论存在 IND-CSA 攻击者的情况，即攻击者未得到加密密钥和隐写密钥时的情形。

在实验 $\mathrm{Exp}^{\text{IND-CSA}}_{\{\Phi',\Phi\}}(A,(\lambda,\lambda'))$ 的第 1 阶段，挑战者利用加密参数 λ 生成加密密钥 p，用隐写参数 λ' 生成隐写密钥 p'；攻击者可任意多次访问喻示 O_1，并从消息集合 $\boldsymbol{M}$ 中任意选取待嵌信息 M。在第 2 阶段，挑战者从载体集合 $\boldsymbol{C}$ 中任意选取载体 C，并随机选择 1 bit b，若 $b=1$，则挑战者在 C 中嵌入消息 M 得到 S，利用 p' 对 S 进行置乱得到 $\mathrm{Per}(S)$，并加密 $\mathrm{Per}(S)$ 得到 C^*，并将 C^* 返回给攻击者；若 $b=0$，利用 p' 对 C 进行置乱得到 $\mathrm{Per}(C)$，利用加密算法将 $\mathrm{Per}(C)$ 加密得到 C^*，将 C^* 返回给攻击者。攻击者可结合 M 和 C^*，任意多次访问喻示 O_2，并判断 C^* 是否载密，并将判断结果 b' 返回给挑战者。其中，若判断 C^* 载密，$b'=1$；否则，返回 $b'=0$。若攻击者判断正确，则 $\mathrm{Exp}^{\text{IND-CSA}}_{\{\Phi,\Phi'\}}(A,(\lambda,\lambda'))$ 取值为 1，否则取值为 0。

在未知加密密钥情况下，根据安全加密体制的 IND-CPA 假设[13]，攻击者在 $\mathrm{Exp}^{\text{IND-CSA}}_{\{\Phi,\Phi'\}}(A,(\lambda,\lambda'))$ 中

无法有效区分 $H(C^*)$ 是由某一载体数据加密生成的密文还是另一载体数据加密生成的密文。记攻击者正确解密 C^* 的概率 $\Pr(\text{Dec\_Success})=\varepsilon$，则解密 C^* 失败的概率 $\Pr(\text{Dec\_Fail})=1-\varepsilon$。若攻击者对密文解密失败，则由于其无法还原载体，故无法进行隐写检测。

对 IND-CSA 攻击者而言，有

$$\text{Adv}_{\{\Phi,\Phi'\}}^{\text{IND-CSA}}(A,(\lambda,\lambda'))\le 2\left|\frac{1+\varepsilon'}{2}\varepsilon+\frac{1}{2}(1-\varepsilon)-\frac{1}{2}\right|=\varepsilon\varepsilon'$$

故在 IND-CSA 条件下，当 $\Pr(\text{Dec\_Success})=\varepsilon$ 且 Φ' 为 ε'-安全时，$\{\Phi',\Phi\}$ 为 $\varepsilon\varepsilon'$-安全的。

综上所述，本文算法在密文域和明文域均能有效抵抗隐写检测分析。在通信双方被强制要求强制解密、加密密钥泄漏和密钥未泄露 3 种情形下，本文的密文域隐写算法都具有较高的安全性，且在上述 3 种攻击条件下，算法的安全性依次增强。

4 结束语

本文结合双层隐写技术和同态加密技术设计了一种 P2C 模式的密文域图像隐写算法。理论分析和实验结果表明该隐写方法在明文域和密文域均具有较高的抗统计检测性能。本文的研究结果不仅可用于隐蔽通信，还适用于密文域数据标注，可为云计算环境下的密文域数据管理和检索提供思路。

参 考 文 献

[1] 冯登国, 张敏, 张妍, 等. 云计算安全研究[J]. 软件学报, 2011, 22(1): 71–83.
Feng Deng-guo, Zhang Min, Zhang Yan, *et al*.. Study on cloud computing security[J]. *Journal of Software*, 2011, 22(1): 71–83.

[2] Gentry C. Fully homomorphic encryption using ideal lattices[C]. Proceedings of the 2009 ACM Symposium on Theory of Computing, STOC'09, NY, USA, 2009: 169–178.

[3] Dijk M V, Gentry C, Halevi S, *et al*.. Fully homomorphic encryption using ideal lattices[C]. Proceedings of in Advances in Cryptology – EUROCRYPT, Riviera, French, 2010, LNCS 6110: 24–43.

[4] Cancellaro M, Battisti F, Carli M, *et al*.. A commutative digital image watermarking and encryption method in the tree structured haar transform domain[J]. *Signal Processing: Image Communication*, 2011, 26(1): 1–12.

[5] Kundur D and Karthik K. Video fingerprinting and encryption principles for digital rights management[J]. *IEEE Multimedia*, 2004, 92(6): 918–932.

[6] Memon N and Wong P W. A buyer-seller watermarking protocol[J]. *IEEE Transations on Image Processing*, 2001, 10(4): 643–649.

[7] Zhang X. Reversible data hiding in encrypted image[J]. *IEEE Signal Processing Letters*, 2011, 18(4): 255–258.

[8] Sun Y and Zhang Xing. A kind of covert channel analysis method based on trusted pipeline[C]. 2011 International Conference on Electrical and Control Engineering (ICECE), Yichang, China, 2011: 5660–5663.

[9] Zhang X, Zhang W, and Wang S. Efficient double-layered steganographic embedding[J]. *IET Electronics Letters*, 2007, 43(8): 482–483.

[10] Filler T, Judas J, and Fridrich J. Minimizing embedding impact in steganography using trellis-coded quantization[C]. Proceedings of SPIE, Electronic Imaging, Media Forensics and Security XII, San Jose, CA, USA, 2009, (5): 1–14.

[11] Fridrich J, Goljan M, and Soukal D. Efficient wet paper codes[C]. Proceedings of 7th International Workshop on Information Hiding, Barcelona, Spain, 2005, LNCS 3727: 204–218.

[12] Barbier J and Alt S. Practical insecurity for effective steganalysis[C]. Proceedings of Information Hiding, Santa Barbara, CA, USA, 2008: 195–208.

[13] Hopper N J. Toward a theory of steganography[D]. [Ph.D. dissertation], Carnegie Mellon University, Pittsburgh, PA, USA, 2004: 20–31.

[14] Schaefer G and Stich M. UCID — an uncompressed color image database[C]. Proceedings of SPIE, Storage and Retrieval Methods and Applications for Multimedia, San Jose, California, USA, 2004, 5307: 472–480.

[15] Luo W, Huang F, and Huang J. Edge adaptive image steganography based on LSB matching revisited[J]. *IEEE Transactions on Information Forensics and Security*, 2010, 5(2): 201–214.

[16] Lu Y, Li X, and Yang B. A secure steganography: noisy region embedding[C]. 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing(IIH-MSP 2009), Kyoto, Japan, 2009: 1046–1051.

[17] Cancelli G, Doerr G, Cox I J, *et al*.. Detection of ±1 LSB steganography based on the amplitude of histogram local extrema[C]. Proccedings of 15th IEEE International Conference on Image Processing, San Diego, CA, 2008: 1288–1291.

[18] Cai K, Li X, Zeng T, *et al*.. Reliable histogram features for detecting LSB matching[C]. Proceedings of IEEE 17th International Conference on Image Processing, Hong Kong, 2010: 1761–1764.

[19] 王朔中, 张开文, 张新鹏. 数字密写和密写分析[M]. 清华大学出版社, 2005: 20–22.
Wang Shuo-zhong, Zhang Kai-wen, and Zhang Xin-peng. Steganography and Steganalysis[M]. Tsinghua University Press, 2005: 20–22.

陈嘉勇： 男，1982 年生，博士生，研究方向为网络安全.

王 超： 男，1985 年生，硕士生，研究方向为信息隐藏.

张卫明： 男，1976 年生，副教授，研究方向为密码学和信息隐藏.