# 抗统计分析的 LSB 密写方案

张新鹏 王朔中 张开文
(上海大学通信与信息工程学院,上海 200072)

**摘 要** 由于通过 RS 统计分析和 Chi-square 统计分析,可以察觉以 LSB 方法密写的秘密信息的存在,因此为提高密写方案的安全性,提出了一种改进的 LSB 密写方案,即如果被嵌入的秘密比特与原始灰度的最低位相同,便不作改动;否则根据周围像素作增 1 或减 1 的调整,而在接收方,只需将载体图象的最低位取出即可恢复秘密信息. 大量图象的模拟实验结果说明,该方案不仅可以抵抗 RS 分析和 Chi-square 分析,而且不增加失真度,并可保持计算量小、提取方便的优点. 并表明改进方案对抵抗这两种密写分析是有效的.

**关键词** 数据安全与计算机安全(520·1060) 密写 密写分析 信息隐藏 最低比特位
**中图法分类号**: TP309.7 TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2003)09-1055-06

# A Novel LSB Steganography Scheme Against Statistical Analysis

ZHANG Xin-peng, WANG Shuo-zhong, and ZHANG Kai-wen
(*School of Communication and Information Engineering, Shanghai University, Shanghai* 200072)

**Abstract** RS and Chi-square analyses are powerful steganalytic techniques for detecting the presence of secret information embedded in the least-significant-bit(LSB) plane of digital images. This paper describes an improved LSB steganographic scheme that can effectively resist both RS and Chi-square analyses without sacrificing perceptual quality of the cover image. In the described steganographic technique, a pair of mutually complementary mappings, $F_1$ and $F_{-1}$, is used, leading to a balanced behavior of several statistical parameters explored by the two steganalytic approaches, therefore improved security. In other words, not only can the flipping between $2i$ and $2i+1$ be used, but also changes from $2i$ to $2i-1$ or from $2i-1$ to $2i$ are allowable. Although modification to the host pixels may affect the higher bit planes, the induced distortion to the host image is not increased compared to the simple LSB replacement technique. Extraction of the embedded information can still be accomplished by extracting the LSB plane as in the straight LSB method. The proposed method has low computation complexity and is easy to implement. Analysis of the effectiveness and experimental results are presented.

**Keywords** Steganography, Steganalysis, Information hiding, Least Significant Bit(LSB)

## 0 引 言

密写(Steganography)与数字水印(Digital watermarking)是信息隐藏(Information hiding)的两大分支[1],自 20 世纪 90 年代中期以来,得到了迅速的发展. 密写与数字水印在实现技术上有很多相同之处,但又各有特点,如数字水印的目的是对多媒体产品进行版权保护,更注重秘密信息的稳健性;而密写的目的则是为了将信息秘密、安全地发送出去,由于要尽可能不引起第三方的怀疑,因而更注重发送信息的隐蔽性.

随着社会的发展,信息安全的意义越来越重大. 尤其自美国"9·11"事件以来,国际上反恐怖的呼声越来越高,察觉、控制非法分子在公众传媒中传送秘密信息——密写分析(Steganalysis)——更受到了广泛关注. 学术界对密写分析的研究也正在不断深入.

对密写的攻击可分为被动攻击和主动攻击两类. 其中被动攻击是指察觉数字媒体中秘密信息的存在[2~5];而主动攻击则是指在不过分影响感觉的情况下,对数字媒体进行干扰,使得发送方隐藏的秘密信息无法被接收方提取[6]. 在实际应用中,只要能

基金项目:国家自然科学基金项目(60072030)和上海市重点学科建设项目(2001-41)
收稿日期:2002-12-16;改回日期:2003-04-07

检测出载体数据中含有秘密信息,那么密写攻击就已经成功.就目前来说,关于被动攻击的研究要多于对主动攻击的研究,但值得注意的是,如果密写方估计到通信网络的管理者会利用统计工具对传输的数据内容进行分析,并有可能将可疑数据滤除,进而追查发送者,那么,密写方可以在嵌入秘密信息后,再对含有秘密信息的数据进行处理,使其统计特性符合正常标准,以瞒过网络管理者[7].另外,如果密写方知晓分析算法的细节,也可以设计出能够抵抗该分析算法的密写方案[8].

LSB(Least Significant Bit)密写方法出现较早,这是一种用秘密信息替换载体数据最不重要比特位的方法.虽然这种方法的稳健性不是很好,但由于其实现简单、隐藏数据大,因此仍然是一种很有希望的密写方法.RS(Regular-Singular)方法就是利用了LSB 方法只改变最不重要比特位的特点,因而可以对载体进行统计分析.这种方法不但能察觉秘密信息的存在,而且可以估计出隐藏的信息量[2,3].Chi-square 统计法则是利用了载体数据的直方图特性来进行分析,在某些情况下,它可以检测出载体中含有秘密信息的区域和嵌入量[4,5].本文对原始的 LSB 密写方法进行了发展,提出了一种可以抵抗 RS 统计分析和 Chi-square 统计分析的改进方案,其对推动密写和密写分析有一定的意义.

## 1 LSB 密写分析

### 1.1 RS 密写分析方法

RS 密写分析方法是由 Fridrich 等提出的[2,3].大家知道,对于绝大多数图象,采样点之间是具有较强相关性的,而秘密信息由于通常经过压缩或加密,可以认为不具有相关性,所以,当秘密信息被嵌入到载体图象数据的最低位后,像素灰度值之间的相关性会在一定程度上受到破坏.RS 方法就是利用这个特性来检测数字媒体中是否含有秘密信息.具体步骤如下:

给定一个图象块,可以用下式表示混乱程度

$$f(\boldsymbol{X}) = \sum |\boldsymbol{X} - \boldsymbol{X}_1| + \sum |\boldsymbol{X} - \boldsymbol{X}_2| \quad (1)$$

其中,$\boldsymbol{X}$ 是图象块的灰度值矩阵,$\boldsymbol{X}_1$ 表示将 $\boldsymbol{X}$ 左移一列,$\boldsymbol{X}_2$ 表示将 $\boldsymbol{X}$ 下移一行,$f(\boldsymbol{X})$表示相邻像素灰度差值的绝对值总和.$f(\boldsymbol{X})$越大,表示该图象块越混乱、相邻像素间的相关性越小.记 $F_1$ 为 $2i$ 与 $2i+1$ 的互相映射关系,即 $0 \leftrightarrow 1, 2 \leftrightarrow 3, \cdots, 254 \leftrightarrow 255$;$F_{-1}$为 $2i-1$ 与 $2i$ 的互相映射关系,即 $-1 \leftrightarrow 0, 1 \leftrightarrow 2, \cdots, 255 \leftrightarrow 256$.LSB 密写就相当于对部分像素应用 $F_1$ 映射.

RS 密写分析时,首先将待检图象分为很多大小相等的图象块,再对每个小图象块随机抽取部分(如 1/2)像素进行 $F_1$ 映射,然后利用式(1)计算其混乱程度是否增加,并计算混乱度增加的图象块在所有图象块中占的比例,记为 $R_+$;而将混乱度减小的图象块在所有图象块中占的比例记为 $S_+$,一般说来 $R_+ + S_+ < 1$;最后应用 $F_{-1}$映射在每个小图象块中进行类似的处理,也记下混乱度增加和减小的图象块的比例,分别为 $R_-$、$S_-$.

如果待检图象没有经过 LSB 密写处理,那么无论应用 $F_1$ 映射,还是应用 $F_{-1}$映射,从统计特性上来说,会同等程度地增加图象块的混乱度,也就是说,$R_+ \approx R_-$、$S_+ \approx S_-$,而且 $R_+ > S_+$、$R_- > S_-$.

如果待检图象是经过 LSB 密写处理的,则无论应用 $F_1$ 映射,还是应用 $F_{-1}$映射的结果都会有明显差别.具体地说,对原始载体图象进行密写本来就是对部分像素应用了 $F_1$ 映射,现在,再对密写图象的部分像素应用 $F_1$ 映射,这样所有像素就可以分为没有被映射处理的、经历过一次映射的、经历过二次映射的 3 类,其中第 3 类像素经历了两次 $F_1$ 映射,又回到了原始值;而如果对密写图象的部分像素应用 $F_{-1}$映射,也会有一些像素经历了两次映射,但由于这些像素经历的是一次 $F_1$ 映射和一次 $F_{-1}$映射,与原始值就会偏离得更远,因此,应用 $F_{-1}$映射对混乱度的增加要大于应用 $F_1$ 映射对混乱度的增加,就会有 $R_+ < R_-$的结果,也就是说,$R_- - S_- > R_+ - S_+$.

RS 密写分析法首先通过计算 $R_+$、$S_+$、$R_-$、$S_-$,然后通过比较它们的关系来检测载体图象数据中是否含有秘密信息.另外,RS 统计还可以进一步对秘密信息量进行估计,即先根据载体图象计算 $R_+$、$S_+$、$R_-$、$S_-$,然后将载体图象数据最低比特全部翻转后,再使用同样的方法计算 $\tilde{R}_+$、$\tilde{S}_+$、$\tilde{R}_-$、$\tilde{S}_-$;最后根据这两组数据估计出秘密信息量,具体细节在这里就不讨论了.

### 1.2 Chi-square 方法[4,5]

假设载体为 8bit 灰度图象,灰度值为 $j$ 的像素数为 $n_j$.在 Chi-square 方法中,是将 $n_{2i}$与 $n_{2i+1}$作为一对数字来处理的,其秘密信息可以看作是 0、1 随机分布的比特流,而且值为 0 与值为 1 的可能性都是1/2.如果秘密信息完全替代了载体图象的最低

位,那么 $n_{2i}$、$n_{2i+1}$的值会比较接近.如果载体图象未经密写,而 $n_{2i}$、$n_{2i+1}$的值则会相差得远一些.定量分析方法如下:令

$$\tilde{n}_{2i}=\frac{n_{2i}+n_{2i+1}}{2} \tag{2}$$

计算

$$\chi^2=\sum_{i=1}^{k}\frac{(n_{2i}-\tilde{n}_{2i})^2}{\tilde{n}_{2i}} \tag{3}$$

其中,$k$ 为由 $n_{2i}$、$n_{2i+1}$组成的数字对的数目.由 $\chi^2$ 计算

$$p=1-\frac{1}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_0^{\chi^2}\exp\left(-\frac{t}{2}\right)t^{\frac{k-1}{2}}\mathrm{d}t \tag{4}$$

$\Gamma(\cdot)$为伽马函数,$p$ 表示载体被密写的概率,如果 $p$ 接近于 100%,则说明载体图象中含有秘密信息.

## 2　抗密写分析的 LSB 改进方案

在 LSB 密写方案中,是按如下方法进行密写的,如果秘密信息位与隐藏该位的像素灰度值的最后一位相同,就不改变原始载体;反之,则要改变灰度值的最后一位,即将 $2i$ 改为 $2i+1$ 或将 $2i+1$ 改为 $2i$,而不会将 $2i$ 改为 $2i-1$ 或将 $2i+1$ 改为 $2i+2$.RS 方法和 Chi-square 方法都是利用 LSB 密写的这个特性进行密写分析的.在 RS 方法中,$F_1$ 映射与 $F_{-1}$映射是具有互补性的,而当密写操作改变原始灰度值时,由于其相当于只应用 $F_1$ 映射,而不会应用 $F_{-1}$映射,所以,RS 方法将 $F_{-1}$引入检验,就可以觉察密写行为.Chi-square 方法则是通过检测灰度值为 $2i$ 和 $2i+1$ 的像素数的接近程度来判断载体是否含有秘密信息.为了抵抗 RS 分析和 Chi-square 分析,可以对 LSB 方法进行改进,即在信息隐藏时,不但可将 $2i$ 改为 $2i+1$ 或将 $2i+1$ 改为 $2i$,也可将 $2i$ 改为 $2i-1$ 或将 $2i+1$ 改为 $2i+2$.具体方法如下.

### 2.1　密写方案

设秘密信息位为 $w$,对应隐藏该秘密信息位的像素灰度值为 $x_{i,j}$.

如果 $w$ 与 $x_{i,j}$的最后一位相同,即 $w=x_{i,j}\bmod 2$,于是就不改变原始数据.

当 $w$ 与 $x_{i,j}$的最后一位不同,即 $w\neq x_{i,j}\bmod 2$ 时,则计算

$$T=\sum_{u=i-1}^{i+1}\sum_{v=j-1}^{j+1}x_{u,v}-9\cdot x_{i,j} \tag{5}$$

对 $x_{i,j}$作如下调整

$$x_{i,j}=\begin{cases}x_{i,j}-1 & \text{if}\quad T\leq 0,0<x_{i,j}<255\\ x_{i,j}+1 & \text{if}\quad T>0,0<x_{i,j}<255\\ x_{i,j}-1 & \text{if}\quad x_{i,j}=255\\ x_{i,j}+1 & \text{if}\quad x_{i,j}=0\end{cases} \tag{6}$$

在式(6)中,无论将 $x_{i,j}$增 1 还是减 1,其最后一位都会变化.根据 $T$ 确定 $x_{i,j}$增减的目的是为了使密写不过分影响相邻像素之间的相关性.

按式(6)进行修改,不仅仅会改变最低比特位,也可能影响多个比特位.尽管这样,改进方案与原有的 LSB 密写技术相比,由于修改的幅度并没有增大,所以失真度还是相同的,而且这种改进方案对密写信息的提取是简单的,即只要将含有秘密信息的像素灰度值的最后一位取出即可.

### 2.2　抗 RS 分析

应用改进方案对载体进行密写时,会有大约一半像素的最低比特位因与秘密信息相同而不发生变化,而其余大约一半的像素的灰度值则会发生变化,而且在这灰度值发生变化的像素中,又会有大约一半的像素灰度值由 $2i$ 变为 $2i+1$ 或由 $2i+1$ 变为 $2i$,即相当于应用了 $F_1$ 映射;而另外大约一半的像素灰度值则由 $2i$ 变为 $2i-1$ 或由 $2i+1$ 变为 $2i+2$,即相当于应用了 $F_{-1}$映射.当使用 RS 统计法进行分析时,因为无论应用 $F_1$ 映射,还是应用 $F_{-1}$映射都会以同样的程度增加图象的混乱度,所以 $R_+$ 与 $R_-$ 很近似,$S_+$ 与 $S$ 也很近似;而对不经密写的图象进行统计,由于也会得到 $R_+\approx R_-$、$S_+\approx S$ 的结果.所以就无法判断是否含有秘密信息.

下面考虑次低位面的情况.最低比特与秘密信息相同时或应用 $F_1$ 映射时,一般像素灰度的次低比特位不会变化,但由于改进方案在大约四分之一的像素上应用了 $F_{-1}$映射,因此这些像素灰度的次低比特位则会发生变化.在这些像素中,又有大约一半只改变次低比特位,另一半像素的倒数第 3 位也会发生变化,即在低位面上同样是一部分应用了 $F_1$ 映射,另一部分应用了 $F_{-1}$映射,所以与最低比特位类似,在次低位面上应用 RS 统计分析也不会奏效.同理,在其他位面上应用 RS 统计分析也无法检测出秘密信息是否存在.

### 2.3　抗 Chi-square 分析

用原有的 LSB 方法进行密写时,由于灰度值为 $2i$ 的像素中会有大约一半改为 $2i+1$,灰度值为 $2i+1$ 的像素中也会有大约一半改为 $2i$,因此,在密写后的载体中,$n_{2i}$、$n_{2i+1}$的值会比较接近.如果采用改

进方案，灰度值为 $j$ 的像素中会有大约一半不变，大约四分之一变为 $j+1$，剩余大约四分之一变为 $j-1$. 设原始载体数据中灰度值为 $j$ 的像素为 $m_j$，密写后的载体中灰度值为 $j$ 的像素为 $n_j$，则有

$$E(n_j)=\frac{m_{j-1}}{4}+\frac{m_j}{2}+\frac{m_{j+1}}{4} \tag{7}$$

其中，$E(\cdot)$表示期望. 根据式(7)可得

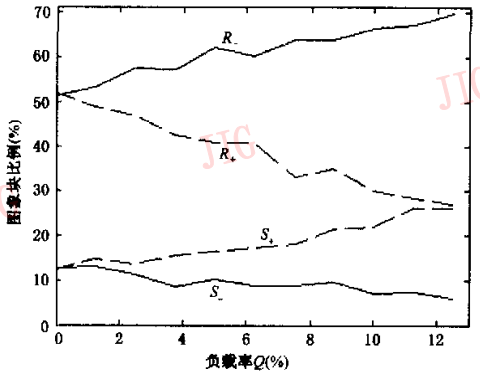
$$E(n_{2i})=\frac{m_{2i-1}}{4}+\frac{m_{2i}}{2}+\frac{m_{2i+1}}{4} \tag{8}$$

$$E(n_{2i+1})=\frac{m_{2i}}{4}+\frac{m_{2i+1}}{2}+\frac{m_{2i+2}}{4} \tag{9}$$

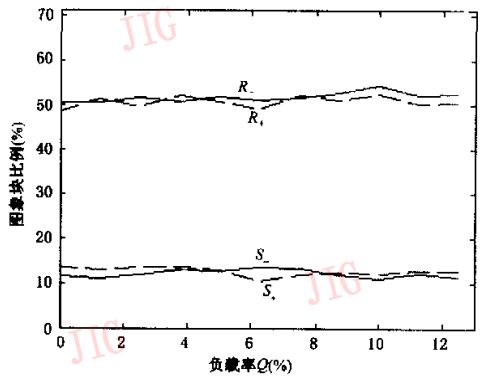
一般情况下，由于 $E(n_{2i})\neq E(n_{2i+1})$，所以应用 Chi-square 方法进行分析时，无论载体是否含有秘密信息，$p$ 总是接近于 0.

## 3 实验结果

对 512×512pixels 的 8bit 标准灰度图象 Lena 分别使用原有的 LSB 技术和改进方案进行了密写对比实验，然后用 RS 方法进行统计特性分析，以查明两种方案的差别. 其结果如图 1 所示，横坐标表示秘密信息负载率 $Q$，即秘密信息比特数除以载体图象大小再除以 8，纵坐标表示 $R_+$、$R_-$、$S_+$、$S_-$ 的值，即图象块比例；图中上方的实线代表 $R_-$，虚线代表 $R_+$，下方的虚线代表 $S_+$，实线代表 $S_-$.




(a) 对使用原有的 LSB 技术密写的分析




(b) 对改进方案密写的分析

图 1 密写图象 Lena 的 RS 统计分析结果

从图 1 可以看出，RS 方法可以对原有的 LSB 技术进行可靠检测，而且信息负载率越大时，$R_+$ 与 $R_-$ 的差距越大、检测性能越好；而对于改进的 LSB 密写方案，当信息负载率提高时，$R_+$、$R_-$、$S_+$、$S_-$ 与不含秘密信息时并没有明显区别，由此可见，对改进的 LSB 密写方案，RS 方法不能检测出秘密信息的存在.

实验时，首先用数码相机采集 100 幅大小为 300×400、灰度值为 8bit 的含风景和人物的图象，然后对这些图象分别计算不含秘密信息、使用原有 LSB 技术密写和使用改进方案密写时的 $R_+-R_-$ 值，结果如图 2 所示. 当对原始图象进行测试时，因为 $R_+$ 与 $R_-$ 很接近，所以 $R_--R_+$ 曲线在 0 附近波动. 当采用原有 LSB 技术进行负载率为 10%的密写后，由于会有 $R_+<R_-$ 的结果，因此用 RS 方法就可以根据这样的结果检测到秘密信息的存在. 而使用改进方案进行负载率 $Q=10\%$ 的密写后，因为 $R_+$ 与 $R_-$ 依然很接近，所以 RS 方法对改进的 LSB 密写方案是无法进行有效检测的.

同时还对多幅图象进行了 Chi-square 分析结果的比较，图 3 给出了对 3 幅标准灰度图象进行测试的情况，即首先应用原有的 LSB 技术对载体图象位于左上角的一半像素进行密写，然后作 Chi-square 统计分析，其结果如图 3(a)所示. 作 Chi-square 分析时，先对载体图象左上角的小部分像素进行处理，并计算 $p$；然后逐渐增加被处理的像素数，直至分析区域扩大至整幅图象. 图 3(a)中横坐标表示分析区域占整幅图象的比例，纵坐标即计算结果 $p$. 在图 3(a)的左侧，$p$ 接近于 100%(在分析区域小于 10%时，可能会因统计数据过少而发生剧烈起伏)，而在分析区域达到 50%后，$p$ 从 100%降到 0 附近；据此可以估计出秘密信息嵌入的位置和嵌入量. 如果对载体图象的所有像素进行密写，然后作 Chi-square 统计分析，那么 $p$ 始终接近于 100%.

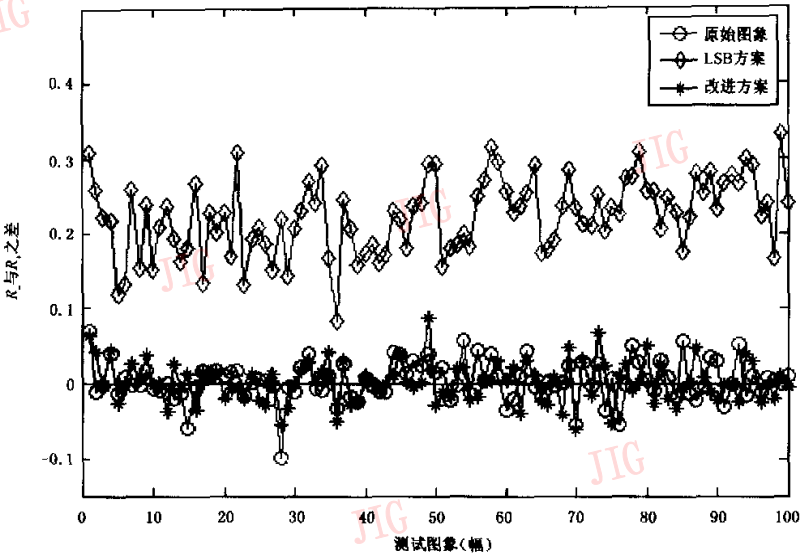
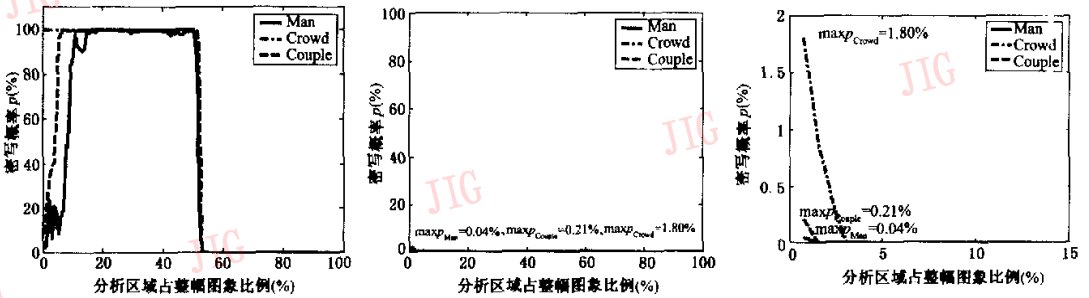



图 2　使用原有的 LSB 技术、使用改进方案以及不含密写信息时 100 幅图象的 $R_- - R_+$ 结果




(a) 对使用原有的 LSB 技术密写的分析　(b) 对使用改进方案密写的分析　(c) 图 3(b)左下角的局部放大

图 3　对采用不同技术密写的 3 幅图象进行 Chi-square 统计分析的结果

使用改进方案对载体图象中位于左上角的一半像素进行密写，Chi-square 分析的结果如图 3(b)所示. 图 3(b)中的 $p$ 值始终接近于 0，3 幅图象密写概率 $p$ 的最大值也仅为 0.04%、1.80%、0.21%(放大情况见图 3(c)). 可见，Chi-square 方法无法对改进方案进行有效的检测.

## 4　结　论

密写与密写分析这两项技术是互相促进、共同发展的. 尽管利用 RS 和 Chi-square 方法能对 LSB 密写进行统计分析，但如果在原有 LSB 方案基础上进行改进，便可以有效地抵御这两种分析方法，并且还能保持计算量小、提取方便的优点. 如何分析这种改进的 LSB 密写方案，将是下一步的研究内容.

## 参 考 文 献

1　Petitcolas F A P, Anderson R J, Kuhn M G. Information Hiding—A Survey[J]. Proceedings of the IEEE, 1999, 87(7): 1062～1078.

2　Fridrich J, Goljan M. Practical steganalysis of digital images—State of the art[A]. In: Proceedings of SPIE on Security and Watermarking of Multimedia Contents IV[C], San Jose, CA, USA, Jan. 2002, 4675: 1～13.

3　Fridrich J, Goljan M, Du R. Detecting LSB steganography in color and gray-scale images[J]. Magazine of IEEE Multimedia, Special Issue on Security. 2001, 9(4): 22～28.

4　Westfeld A, Pfitzmann A. Attacks on steganographic systems [A]. In: Lecture Notes in Computer Science on Third International Workshop on Information Hiding [C], Berlin

Heidelberg German: Springer-Verlag. 1999, 1768: 61～76.

5 Provos N, Honeyman P. Detecting Steganographic Content on the Internet [R]. Center for Information Technology Integration, University of Michigan, USA, Technical Report. 2001: 01～11.

6 Ettinger J. Steganalysis and game equilibria [A]. In: Lecture Notes in Computer Science on Second International Workshop on Information Hiding [C], Berlin Heidelberg German: Springer-Verlag, 1998, 1525: 319～328.

7 Provos N. Defending against statistical steganalysis [A]. In: Proceedings 10th Advanced Computing System Association, Security Symposium [EB/OL]. Washington, DC, USA, 2001. http://www. usenix. org/publications/library/proceedings/sec01/provos. /html

8 Wang S, Zhang X, Zhang K. Steganographic technique capable of withstanding RQP analysis [J]. Journal of Shanghai University (English Edition), 2002, 6(4): 273～277.

**张新鹏** 1975 年生，1995 年获吉林大学数学系理学学士学位，2001 年获上海大学通信与信息工程学院工学硕士学位，现为该校博士研究生. 主要研究领域为信息隐藏、数字水印、数字图象处理等. 已发表论文近 30 篇.

**王朔中** 1943 年生，1966 年毕业于北京大学，1982 年获英国伯明翰大学博士学位，现任上海大学通信与信息工程学院教授. 研究领域为图象处理、音频信号处理、信息隐藏.

**张开文** 1963 年生，1982 年毕业于解放军工程技术学院无线电工程专业，1997 年获得该校信息处理专业硕士学位. 现为副研究员、上海大学通信与信息工程学院博士研究生. 主要研究领域为从事信号与信息处理工作、数字水印与信息隐藏等相关技术.