Exploiting Preserved Statistics for Steganalysis

Rainer Böhme and Andreas Westfeld

Technische Universität Dresden
Institute for System Architecture
01062 Dresden, Germany
{rainer.boehme,westfeld}@inf.tu-dresden.de

Abstract. We introduce a steganalytic method which takes advantage of statistics that were preserved to prevent the chi-square attack. We show that preserving statistics by skipping certain groups of pixels—apart from reducing the maximum payload—does not diminish the ability to recognise steganographic modifications. The effect is quite reverse: The new detection method works more reliably than the chi-square attack, if the same message was embedded by overwriting least significant bits and straddled over the whole image.

1 Introduction

Steganography means "covered writing." Steganographic programs are capable of embedding a message into innocuous looking carrier media. Carrier media can be digitised images sent as E-mail attachments or found in eBay offers. The carrier medium is slightly modified by the embedding function so that an attacker should not perceive such changes. Steganography is one way to communicate confidentially: non-involved persons do not notice whether the secret message exists or not.

If cryptography is used to communicate secretly, a third party may still notice when an encrypted message is sent. However, she cannot read its content. In some countries, such as China, there are legal restrictions for the usage of cryptography [11]. People that are not allowed to encrypt their E-mail may fall back to steganography and embed their secrets in images to transfer them unnoticeable to the receiver.

Beside the topmost goal of changing the carrier medium as inconspicuously as possible, steganographic algorithms try to implement other helpful properties, such as a large payload and an error-free readability of the embedded content after transmission over a distorted channel (e. g., in a radio contact). It is obvious that these are conflicting goals. For example, steganographic changes are less recognisable if payloads keep small.

Apparently, it is hard to satisfy the theoretical security conditions [2,10,16] in practical implementations. Hence, new algorithms are proven to be secure against known attacks and obvious derivations. It is for this reason, that steganalysis, the art of detecting steganographic changes, is so successful in forms and manners [9,15]. Steganalytic attacks aim to detect the use of steganography.

J. Fridrich (Ed.): IH 2004, LNCS 3200, pp. 82–96, 2004.

There is a recurrent alternation of improved embedding methods and successful attacks breaking these. Following this tradition, we analyse a steganographic algorithm that was presented by Franz at the last workshop [3]. She constructed this algorithm to overcome histogram attacks. Her new algorithm is based on an embedding function that overwrites the least significant bits (LSB) of a carrier. The pure application of this method is detectable by visual and statistical chi-square attacks [15]. So, Franz restricts the embedding function to selected pixels to keep the histogram (first order statistics) together with the image structure (second order statistics). These measures secure the algorithm against the aforementioned attacks.[1]

This paper is structured as follows: In the next section we describe the embedding algorithm proposed in [3], which was designed to preserve statistical properties (PSP) of the carrier image. This algorithm basically extends the method of overwriting the least significant bits (LSB) to prevent chi-square attacks presented in [15]. Then, in Sect. 3, we outline an attacking strategy which exploits the preserved statistics. As the embedding algorithm keeps some relevant distributions in the co-occurrence matrix, an attacker can reproduce the classification criteria applied while embedding. A comparison between the resulting two sets of usable and unusable pixels reveals typical embedding artefacts of the PSP method, which is therefore detectable. Our experimental results (see Sect. 4) indicate that the proposed attack detects PSP steganography even more reliably than the chi-square attack does on simple LSB embedded data of comparable capacity. In Sect. 5, we describe possible countermeasures and discuss their impact on capacity and security. A final conclusion for future improvements of steganographic algorithms is given in Sect. 6.

2 "Preserving Statistical Properties" Algorithm

The "Preserving Statistical Properties" (PSP) algorithm is an extension to the widely used method of overwriting the least significant bits (LSB) in digitised media data. Both algorithms, LSB as well as PSP, embed steganographic messages into the spatial domain representation of uncompressed or losslessly compressed image data. Given a $X \times Y$ sized greyscale image $B = \{0, \ldots, N-1\}^{X,Y}$ with N possible shades, let

$$S_k = \{(x, y) | b_{x,y} = k\},\ 0 \leq x < X,\ 0 \leq y < Y,\ 0 \leq k < N$$

be the set of pixels in B with shade k. Obviously the shades S_k are disjoint with each other. Both algorithms assume that the shades $S_{0,\ldots,N-1}$ can be grouped into $\frac{N}{2^l}$ groups G of 2^l shades ($l = 1, 2, \ldots$), so that a replacement with any member of the same group is imperceptible. Let $\mathcal{G}$ be the set of all groups G in a given image. The information which shade $b_{x,y}$ of the visually indistinguishable group members actually occurs at a certain position (x, y) can be used for

[1] As recent analyses showed vulnerable cases against the RS attack [7], Franz addresses the problem that the new method does not consider all higher order statistics [5].

steganographic messages. Grouping shades that differ only in the least significant bit, is the most common way to fulfil this assumption. This leads to $|\mathcal{G}| = N/2$ groups

$$G_k = S_{2k} \cup S_{2k+1},\ 0 \leq k < |\mathcal{G}|,$$

and a maximum steganographic capacity of one bit per pixel. The imperceptibility assumption is plausible for the least significant bit, because adjacent shades differ minimum in brightness and are at most exposed to quantisation noise. Further generalisations, e. g., colour components of true colour images or indices in sorted palette entries, are extraneous to the following considerations and we therefore forgo a detailed discussion.

The presented LSB method is known to be vulnerable against the chi-square attack presented in [15]. Overwriting the least significant bits according to a uniform distributed message equalises the individual within-group distributions. These pair wise adjustments can be reliably detected by a chi-square goodness-of-fit test between the empirical distributions of $|S_{2k}|$, and $|S_{2k+1}|$, respectively, against the expected distribution for a maximum embedded message

$$\frac{|S_{2k}| + |S_{2k+1}|}{2} = \frac{|G_k|}{2},\ 0 \leq k < |\mathcal{G}|.$$

The PSP algorithm was designed to resist the chi-square attack and introduces two countermeasures, such as classification of groups and skewness corrected embedding. Both measures are adaptive, i. e., they depend on the content of the carrier image, and both reduce the maximum length of the hidden message.

In this paper, we use the term *classification* of groups to describe a preselection process, which distinguishes groups $\mathcal{G}^+ \subset \mathcal{G}$ that are safe for LSB embedding from $\mathcal{G}^- = \mathcal{G} \backslash \mathcal{G}^+$, that are not. The chi-square attack is successful against LSB embedding, because even heavily unequal distributions of group members are equalised during embedding. Typical skewness between group members results from plain surfaces as well as from saturated areas in the carrier image. To preserve these characteristics, within-group dependency tests are run on co-occurrence matrices C for each group G_k. Only those groups $G_k \in \mathcal{G}^+$ that fail the dependency tests are classified as "safe groups" and thus are used for embedding.

A co-occurrence matrix is a transition histogram between adjacent pixels for a defined relation in the spatial domain. It contains the frequency of a certain shade depending on the shade of a defined neighbour. As described in [3], we calculate

$$c_{i,j} = |\{(i,j) | b_{x,y} = i \ \ \wedge \ \ b_{x+\Delta x, y+\Delta y} = j\}|,$$
$$0 \leq i,j < N,\ 0 \leq x < X,\ 0 \leq y < Y$$

for each of the following relations $(\Delta x, \Delta y) \in \{(1,0), (-1,1), (0,1), (1,1)\}$ and test the within-group dependency with four fourfold contingency tables (cf. Table 1). The relevant entries for the dependency calculations are marked boldface in the following co-occurrence matrix

Table 1. Contingency table for classification of group G_k

(x, y)	$(x + \Delta x, y + \Delta y)$		
	$\in S_{2k}$	$\in S_{2k+1}$	$\sum$
$\in S_{2k}$	$c_{2k,2k}$	$c_{2k,2k+1}$	c'_{2k}
$\in S_{2k+1}$	$c_{2k+1,2k}$	$c_{2k+1,2k+1}$	c'_{2k+1}
$\sum$	c''_{2k}	c''_{2k+1}	n

$$C = \begin{pmatrix}
\mathbf{c_{0,0}} & \mathbf{c_{1,0}} & c_{2,0} & c_{3,0} & \cdots & c_{2k,0} & c_{2k+1,0} & \cdots & c_{254,0} & c_{255,0} \\
\mathbf{c_{0,1}} & \mathbf{c_{1,1}} & c_{2,1} & c_{3,1} & \cdots & c_{2k,1} & c_{2k+1,1} & \cdots & c_{254,1} & c_{255,1} \\
c_{0,2} & c_{1,2} & \mathbf{c_{2,2}} & \mathbf{c_{3,2}} & \cdots & c_{2k,2} & c_{2k+1,2} & \cdots & c_{254,2} & c_{255,2} \\
c_{0,3} & c_{1,3} & \mathbf{c_{2,3}} & \mathbf{c_{3,3}} & \cdots & c_{2k,3} & c_{2k+1,3} & \cdots & c_{254,3} & c_{255,3} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
c_{0,2k} & c_{1,2k} & c_{2,2k} & c_{3,2k} & \cdots & \mathbf{c_{2k,2k}} & \mathbf{c_{2k+1,2k}} & \cdots & c_{254,2k} & c_{255,2k} \\
c_{0,2k+1} & c_{1,2k+1} & c_{2,2k+1} & c_{3,2k+1} & \cdots & \mathbf{c_{2k,2k+1}} & \mathbf{c_{2k+1,2k+1}} & \cdots & c_{254,2k+1} & c_{255,2k+1} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
c_{0,254} & c_{1,254} & c_{2,254} & c_{3,254} & \cdots & c_{2k,254} & c_{2k+1,254} & \cdots & \mathbf{c_{254,254}} & \mathbf{c_{255,254}} \\
c_{0,255} & c_{1,255} & c_{2,255} & c_{3,255} & \cdots & c_{2k,255} & c_{2k+1,255} & \cdots & \mathbf{c_{254,255}} & \mathbf{c_{255,255}}
\end{pmatrix}$$

The test statistics χ^2 is calculated according to the following equation

$$\chi^2 = \frac{n(c_{2k,2k}c_{2k+1,2k+1} - c_{2k,2k+1}c_{2k+1,2k})^2}{c'_{2k}\ c'_{2k+1}\ c''_{2k}\ c''_{2k+1}}.$$

We assume independency for values less than $\chi^2 < 3.84$, corresponding to a significance level of $p_\alpha > 0.05$. If one of the four tests rejects the null hypothesis, the whole group is classified as unsafe and excluded from embedding.[4]

For example, 40 shades (15 %) of our example image shown in Fig. 1 were excluded. They cover 29.9 % of the surface and are marked white in Fig. 2. Further examinations with our test database indicate an average share of 43 % of the shades classified as unsafe causing an average loss of 30 % of usable pixels.

As a second modification, the PSP algorithm overwrites the least significant bits with exactly the same distribution as found in the carrier to avoid changes in the first order statistics. This systematic change is the Achilles' heel of LSB embedding and enables successful chi-square attacks with simple histogram analyses. In contrast, PSP makes effort to adopt the message distribution to the prior proportion by adding additional bits of required value and subsequently permuting the message [3]. This second modification limits the capacity of group G_k to $2 \cdot \min(|S_{2k}|, |S_{2k+1}|)$ on average. Assuming a perfectly matching code, the upper bound for the capacity of group G_k can be described with the entropy relation [14]

Fig. 1. Example greyscale image

Fig. 2. Steganographically useable pixels in the example image

$$H_k = -|S_{2k}| \log_2 \frac{|S_{2k}|}{|G_k|} - |S_{2k+1}| \log_2 \frac{|S_{2k+1}|}{|G_k|}.$$

However, the method employed by Franz does not achieve this limit. Using an arithmetic decoding operation, as proposed in [13], offers a more elegant way to preserve first order statistics—but not the exact frequencies—while embedding message bits.

Both measures together, group classification and adaptive message distribution[2], make PSP embedding secure against chi-square attacks (cf. Sect. 4).

Figure 3 contrasts LSB embedding with PSP embedding on a typical gradient part taken from an example image. The white zigzag lines separate shades belonging to different groups. For demonstration purpose, we assume that the shades S_4 and S_5 are excluded from embedding in the PSP case. Also, on the bottom line, the combined co-occurrence matrices are given for the four applied relations

$$(\Delta x, \Delta y) \in \left\{ \begin{matrix} (1,0) & (-1,1) \\ (0,1) & (1,1) \end{matrix} \right\}, \qquad \boxed{\begin{matrix} \rightarrow & \swarrow \\ \downarrow & \searrow \end{matrix}}$$

where *combined* means that the respective elements of the four resulting co-occurrence matrices are printed in each cell.

As the histograms in the middle indicate, the PSP method is not vulnerable to pair wise levelling of shade frequencies: The first order statistics from the carrier histogram are successfully preserved.

3 A Detection Strategy for PSP Steganography

A closer look at the co-occurrence matrices reveals that both embedding schemes leave noticeable traits outside the framed within-group contingency tables. According to the PSP algorithm, groups with high within-group dependencies in the co-occurrence matrix are excluded to prevent a complete erasure of those typical dependencies from the image. In fact, interdependencies in the co-occurrence matrix do not only occur inside the frames. Nevertheless, these within-group dependencies are the only information taken into account for the classification decision.

The PSP scheme does not prevent an attacker from evaluating the between-group dependencies. In addition, the preservation of the first order statistics enables the attacker to re-evaluate the classification decisions and separate used from excluded groups. Strong differences in the higher order statistics between the two classes are a reliable indicator for PSP type steganography.

To construct our attack we need some assumptions about the characteristics of image data. So we state that adjacent pixels correlate strongly, i. e., with high probability they differ only minor in brightness. The majority of dissimilar neighbours of pixels in S_k is expected to be a subset of $S_{k-1} \cup S_{k+1}$. For example, in our test database we found almost 60 % of dissimilar adjacent pixels differing

[2] Meanwhile Franz calls these measures *CCM* and *Histo*, respectively [5].

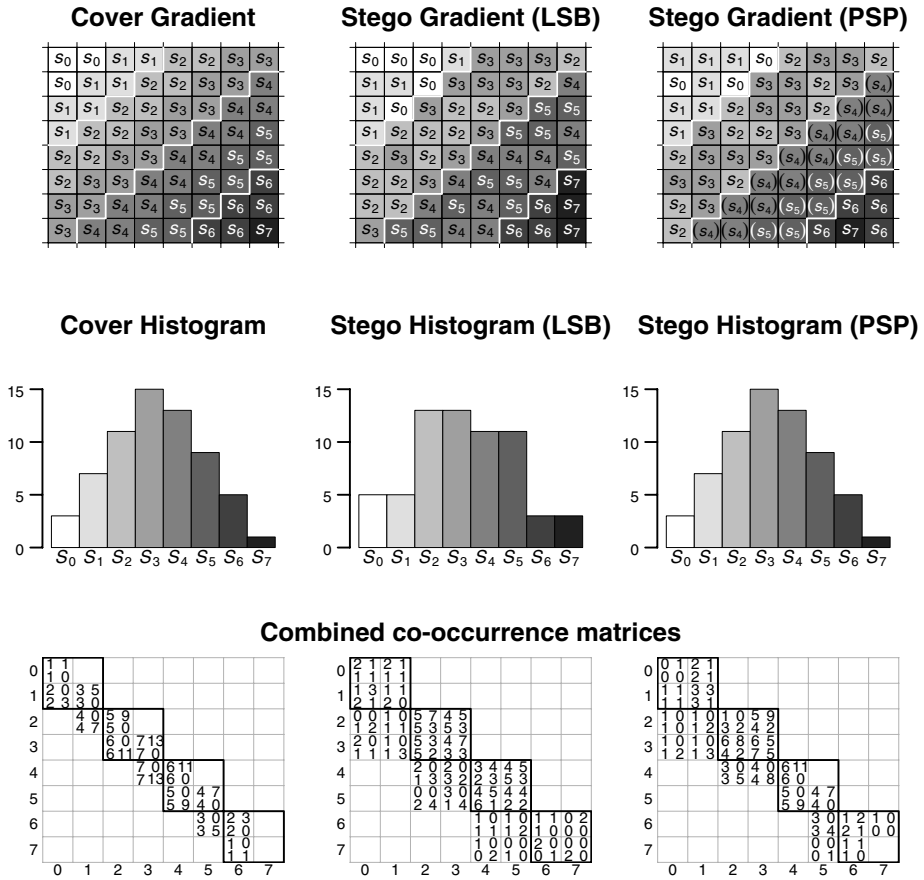

Fig. 3. Comparison of LSB and PSP embedding

by only ±1 in brightness. Hence, any pixel in $S_{<2k}$, darker than the shades in uniformly distributed G_k, is with higher probability neighbour of the darker pixels in $S_{2k} \in G_k$ than the brighter ones in $S_{2k+1} \in G_k$, and vice versa. Still under the assumption that $|S_{2k}| = |S_{2k+1}|$, we assert

$$\begin{aligned} P(b_{x,y} = a | b_{x',y'} = 2k+1) &< P(b_{x,y} = a | b_{x',y'} = 2k) \qquad \text{for } a \leq 2k, \\ P(b_{x,y} = a | b_{x',y'} = 2k+1) &> P(b_{x,y} = a | b_{x',y'} = 2k) \qquad \text{else,} \end{aligned}$$

with $x' = x + \Delta x$, $y' = y + \Delta y$, and $1 \leq \sqrt{\Delta x^2 + \Delta y^2} < 2$. This relation leads to a typical structure in the co-occurrence matrix C. Table 2 shows the two relevant columns for a group G_k and the expected individual proportions between the corresponding frequencies.

Table 2. Structure of G_k columns before embedding

(x', y')	(x, y)		
	$\in S_{2k}$		$\in S_{2k+1}$
$\in S_0$	$c_{2k,0}$	$>$	$c_{2k+1,0}$
$\in S_1$	$c_{2k,1}$	$>$	$c_{2k+1,1}$
$\vdots$	$\vdots$	$>$	$\vdots$
$\in S_{2k-1}$	$c_{2k,2k-1}$	$>$	$c_{2k+1,2k-1}$
$\in S_{2k}$	$c_{2k,2k}$	$>$	$c_{2k+1,2k}$
$\in S_{2k+1}$	$c_{2k,2k+1}$	$<$	$c_{2k+1,2k+1}$
$\in S_{2k+2}$	$c_{2k,2k+2}$	$<$	$c_{2k+1,2k+2}$
$\vdots$	$\vdots$	$<$	$\vdots$
$\in S_{N-2}$	$c_{2k,N-2}$	$<$	$c_{2k+1,N-2}$
$\in S_{N-1}$	$c_{2k,N-1}$	$<$	$c_{2k+1,N-1}$
$\sum$	$\lvert S_{2k}\rvert$	$=$	$\lvert S_{2k+1}\rvert$

As PSP embedding preserves the distribution within the groups and does not mind the neighbourhood relations, it is indistinguishable from a random permutation within each group. Given that $G_k \in \mathcal{G}^+$, it is usable for embedding. The random permutation of the shades within G_k equalises the frequencies for S_{2k} and S_{2k+1} in relation to all other shades in the co-occurrence matrix. The post-embedding structure of the G_k columns in C is shown in Table 3.

Table 3. Structure of G_k columns after PSP embedding

(x', y')	(x, y)		
	$\in S_{2k}$		$\in S_{2k+1}$
$\in S_0$	$c_{2k,0}$	$=$	$c_{2k+1,0}$
$\in S_1$	$c_{2k,1}$	$=$	$c_{2k+1,1}$
$\vdots$	$\vdots$	$=$	$\vdots$
$\in S_{2k-1}$	$c_{2k,2k-1}$	$=$	$c_{2k+1,2k-1}$
$\in S_{2k}$	$c_{2k,2k}$	$=$	$c_{2k+1,2k}$
$\in S_{2k+1}$	$c_{2k,2k+1}$	$=$	$c_{2k+1,2k+1}$
$\in S_{2k+2}$	$c_{2k,2k+2}$	$=$	$c_{2k+1,2k+2}$
$\vdots$	$\vdots$	$=$	$\vdots$
$\in S_{N-2}$	$c_{2k,N-2}$	$=$	$c_{2k+1,N-2}$
$\in S_{N-1}$	$c_{2k,N-1}$	$=$	$c_{2k+1,N-1}$
$\sum$	$\lvert S_{2k}\rvert$	$=$	$\lvert S_{2k+1}\rvert$

We can distinguish the pre- and post-embedding structures shown in Tables 2 and 3 with a contingency test. For this purpose, we interpret a pair of columns

Table 4. G_k columns after embedding with imbalanced frequencies of S_{2k} and S_{2k+1}

(x', y')	(x, y)		
	$\in S_{2k}$		$\in S_{2k+1}$
$\in S_0$	$c_{2k,0}$	$<$	$c_{2k+1,0}$
$\in S_1$	$c_{2k,1}$	$<$	$c_{2k+1,1}$
$\vdots$	$\vdots$	$<$	$\vdots$
$\in S_{2k-1}$	$c_{2k,2k-1}$	$<$	$c_{2k+1,2k-1}$
$\in S_{2k}$	$c_{2k,2k}$	$<$	$c_{2k+1,2k}$
$\in S_{2k+1}$	$c_{2k,2k+1}$	$<$	$c_{2k+1,2k+1}$
$\in S_{2k+2}$	$c_{2k,2k+2}$	$<$	$c_{2k+1,2k+2}$
$\vdots$	$\vdots$	$<$	$\vdots$
$\in S_{N-2}$	$c_{2k,N-2}$	$<$	$c_{2k+1,N-2}$
$\in S_{N-1}$	$c_{2k,N-1}$	$<$	$c_{2k+1,N-1}$
$\sum$	$\lvert S_{2k}\rvert$	$<$	$\lvert S_{2k+1}\rvert$

(x', y')	(x, y)		
	$\in S_{2k}$		$\in S_{2k+1}$
$\in S_0$	$c_{2k,0}$	$>$	$c_{2k+1,0}$
$\in S_1$	$c_{2k,1}$	$>$	$c_{2k+1,1}$
$\vdots$	$\vdots$	$>$	$\vdots$
$\in S_{2k-1}$	$c_{2k,2k-1}$	$>$	$c_{2k+1,2k-1}$
$\in S_{2k}$	$c_{2k,2k}$	$>$	$c_{2k+1,2k}$
$\in S_{2k+1}$	$c_{2k,2k+1}$	$>$	$c_{2k+1,2k+1}$
$\in S_{2k+2}$	$c_{2k,2k+2}$	$>$	$c_{2k+1,2k+2}$
$\vdots$	$\vdots$	$>$	$\vdots$
$\in S_{N-2}$	$c_{2k,N-2}$	$>$	$c_{2k+1,N-2}$
$\in S_{N-1}$	$c_{2k,N-1}$	$>$	$c_{2k+1,N-1}$
$\sum$	$\lvert S_{2k}\rvert$	$>$	$\lvert S_{2k+1}\rvert$

from C as a contingency table and perform a chi-square test for dependency. The former structure is supposed to show a noticeable dependency, the latter not. We further refer to this procedure as *between-group dependency test*.

Even if we drop the assumption that the membership is uniformly distributed between S_{2k} and S_{2k+1} within G_k, we still expect dependencies in the carrier image, modulated by the proportion $S_{2k} : S_{2k+1}$:

$$\begin{aligned} |S_{2k}| \cdot c_{a,2k+1} &< |S_{2k+1}| \cdot c_{a,2k} && \text{for } a < 2k, \\ |S_{2k}| \cdot c_{a,2k+1} &> |S_{2k+1}| \cdot c_{a,2k} && \text{else.} \end{aligned}$$

As the PSP scheme uses adaptive skewness correction, the imbalanced situation is quite probable. Nevertheless, there are still different directions of the inequality relations between adjacent columns of the co-occurrence matrix, which are equally aligned in the groups "permuted" after the PSP embedding operation (cf. Table 4). These alignments are also recognised as independently distributed events by the contingency test. Hence, the skewness correction does not weaken our ability to distinguish between permuted and original groups.

Certain practical obstacles impede using these analyses to guide a precise attack on PSP embedding. At first, the columns of the co-occurrence matrix hold a lot of low frequency entries that bias the outcome of the chi-square between-group dependency test. Secondly, we have to take into account that the above mentioned interrelations apply to all of the four co-occurrence matrices representing the four relations. We tackle these problems by first summing up the four matrices and then erasing rows with row sums less than a minimum count q. All images, whether with or without a PSP embedded message, contain a certain amount of groups that pass the between-group dependency test. Only

the test results of the actually usable groups in $\mathcal{G}^+$ contain valuable information for an attacker about the application of PSP steganography. Therefore, the attacker has to gain knowledge, which groups belong to $\mathcal{G}^+$. Fortunately, this is not difficult, because the PSP scheme preserves the relevant statistics so that the receiver is able to recalculate the initial classification of groups as shown in Sect. 2.

The final step of the proposed attack is an inference from a set of between-group dependency tests to the existence of steganographic content. Since the tests are not accurate for all groups, we cannot expect independency for all members of $\mathcal{G}^+$. Therefore we allow a certain number of tests below a threshold t to pass the between-group dependency test on a $p_\alpha < 0.01$ significance level. It seems sensible to choose q dependent on the number of pixels $X \cdot Y$ and the threshold t on the number of groups $|\mathcal{G}|$. These refinements are subject to further research.

In brief, the attack procedure can be summarised in four steps:

1. Classify all groups according to the embedding scheme,
2. calculate and sum co-occurrence matrices for four relations,
3. test between-group dependencies in column pairs for all usable groups,
4. count positive tests and compare with threshold value.

Our experimental results described in the following section provide a proof of concept for the proposed attack.

4 Experimental Results

To evaluate the practical capabilities of the proposed attack we assembled a test database T_0 of 100 greyscale images sized $X \times Y = 284 \times 213$ pixels ($N = 256$ shades). The images were randomly drawn from a large number of high resolution photographs from a digital camera. An $8:1$ size reduction ensures that possible compression artefacts of the initial JPEG encoding are effectively removed [6]. The small images were stored as losslessly compressed PNG files and analysed with the R software for statistical computing [8,12].

To compare the LSB and PSP embedding schemes, we prepared three test sets:

1. T_1: LSB embedding of uniformly distributed random bits using 100 % of the capacity (i. e., 1 bit per pixel),
2. T_2: PSP embedding of uniformly distributed random bits using 100 % of the capacity (between 0.1 and 1.0 bits per pixel, depending on the image, mean $\mu = 0.77$),
3. T_3: LSB embedding of uniformly distributed random bits using the respective maximum capacity of T_2.

The images of all test sets ($T_0, \ldots, T_3$) were exposed to the chi-square attack with a threshold criteria of $p_\alpha < 0.01$, as well as to the proposed PSP attack

Table 5. Summary of experimental attacks

Attack	Test set, algorithm	Results	
		FALSE	TRUE
Chi-square attack			
	T_0: Plain carrier	92	8
	T_1: LSB (full capacity)	0	100
	T_2: PSP (max capacity)	92	8
	T_3: LSB (limited PSP cap.)	22	78
Proposed attack			
	T_0: Plain carrier	94	6
	T_2: PSP (max capacity)	0	100

Test data: 100 greyscale images sized 284×213 pixel, $N = 256$

with a maximum number of passed tests of $t = 8$, and a minimum row sum of co-occurrence cells $q = 10$. The results are presented in Table 5.

As expected, the chi-square attack reliably identified all LSB steganograms with full capacity usage. However, we noted that eight percent of the tests of pristine material led to a false positive. The same attack applied to the PSP embedded images was comparably ineffective. The preservation of first order statistics successfully prevents chi-square attacks.

Even if invisible to the chi-square attack, all PSP steganograms can be detected with the proposed attack, although the absolute message length is only a fractional amount of the LSB capacity. In fact, four images with less than 20 % of the respective LSB capacity are reliably detected. Regarding the number of false positives, the discriminatory power of the PSP attack seems to exceed the chi-square attack, even though the numbers are too small to provide strong evidence. The tests on T_3 reveal that passing on full capacity and accepting a reduced message length with the well known LSB algorithm is comparatively safer than using the more sophisticated PSP scheme.

To evaluate the stability over different utilisations of capacity between the two embedding schemes with their respective attacks, we gradually reduced the message lengths embedded with the PSP method. In addition, precisely the same amount of bits embedded with PSP was also LSB embedded in the respective images to build a comparison group. As the results in Table 6 indicate, the proposed PSP attack provides higher detection rates for high capacity utilisations.

5 Discussion of Countermeasures

The proposed attack basically exploits the removal of inter-dependencies between adjacent pixels belonging to different groups. A rather naïve approach to tackle

Table 6. Attack reliability against capacity usage

Capacity usage	Embedding density	Attacks	
		chi-square	proposed
% of max. PSP capacity	av. msg. bits per pixel	# of hits out of 100	# of hits out of 100
100 %	0.77	78	100
75 %	0.58	62	88
50 %	0.39	45	38
25 %	0.20	35	14

Test data: 100 greyscale images sized 284×213 pixel, $N = 256$

this problem could be the exclusion of all pixels with neighbours of other groups. So the set of usable pixels will be reduced to those pixels completely surrounded by neighbours of their own group G_k,

$$G'_k = \{(x, y) | (x + \Delta x, y + \Delta y) \in G_k,\ \forall \Delta x, \Delta y \in \{-1, 0, 1\}\}.$$

This modification obviously withstands the proposed attack because the between-group interdependencies are kept untouched. However, only a tiny set of pixels meets this strict condition. For example, our test image contains only 15 usable pixels depicted black in Fig. 4. The comparably larger count of grey pixels in Fig. 4 are also surrounded by the same group but were classified as unsafe according to the PSP classification. Because of the vanishing capacity it is hard to say whether an adapted attack regarding the more distant neighbours ($2 \leq \sqrt{\Delta x^2 + \Delta y^2} < 3$) fails because of the low sample size or is generally impossible. Experiments with larger samples of images with higher resolution are subject to further research.

Regarding this low residual capacity, the LSB algorithm may be a comparably safe alternative. In addition, the security can be further increased by implementing the advices from [15], e. g., to replace LSB overwriting by a more suitable operation such as incrementing or decrementing.

Adaptive embedding, i. e., regarding the carrier structures, is a promising principle for steganography but also opens new pitfalls because the receiver has to recover the structural information to extract the message. For example, the PSP method implements its adaptive mechanism on a group wise classification that can be reproduced both by the receiver but also by an attacker. On the one hand, it is important that the receiver is able to recover all necessary information to extract the message. On the other hand, any information about which pixels are actually usable, gives also an advantage to the attacker: By contrasting the two groups in relevant statistics, she can reveal systematic characteristics that are typical for the embedding scheme but rarely observable in pristine carrier data. We will briefly outline two measures to avoid these problems. At first, the "meta-information approach" aims to hide the classification information by encrypting and embedding it into the safest parts of a carrier.

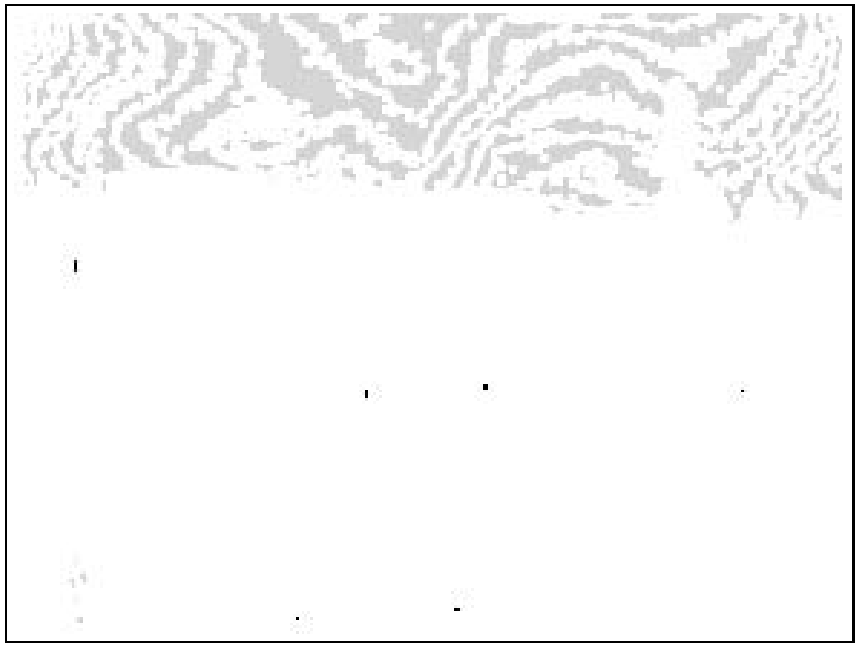

Fig. 4. Pixels surrounded by the same group in our test image. Black pixels belong to safe groups in $\mathcal{G}^+$, grey to unsafe

So, the receiver decodes the meta-information before using it to extract the payload message. Second, the "selection channel approach" [1] completely avoids the share of meta-information concerning the usable positions. Parity encoding ensures that the receiver is always able to extract the message without knowledge about the actually altered bits. Both approaches unfortunately reduce the maximum message length.

6 Conclusion

The presented attack against a steganographic algorithm that preserves some relevant statistics puts into question, whether a rather fussy preservation helps to increase security and therefore should be included in future embedding algorithms. This does not imply that the preservation of statistics is generally a bad idea, but the way it is achieved—i. e., skipping certain "dangerous" groups while modifying others—makes the discussed scheme vulnerable to the proposed attack.

In addition, the exact preservation of statistics that are used for the classification decision enables an attacker to reproduce this decision. This practice causes the sender to give up her superiority of information.

Since first and higher order statistics do heavily vary between different pictures, and given an attacker who has no possibility to guess or estimate these

parameters of the carrier, a moderate change of them does not necessarily weaken security. It may be wise, to refocus further development of steganographic methods from compulsive preservation of parameters to the avoidance of typical—and hence conspicuous—patterns and artefacts. For instance, the promising model-based approach for steganography [13] already employs some of these ideas, even though an adversary can still reproduce the distribution model.

Nevertheless, we suppose that adaptive embedding is a promising practice but classification criteria need to be carefully selected. Using or avoiding shades globally may be problematic in two senses. At first, it raises the danger of misclassifications. For example, a bright shade covering large parts of the sky in our example image also occurs in the lower part. The dependencies in the sky cause a global exclusion of the whole group, even if it could be used for data hiding in the lower part. Vice versa, a shade that is independent at the overwhelming majority of occurrences may be classified as usable even if some occurrences in a "dangerous" context give an attacker strong evidence for steganographic modifications. The second problem of global classification concerns large numbers. The statistical tests of an attacker tend to become the more reliable the more observations she has. Given the situation that a defined message could be transferred either in one large or in several tiny images, we face the following obscurity. With global criteria, the probability of detection increases with the amount of data per pass, i. e., one large image is more dangerous than several tiny images. Therefore, we suggest to research local adaptive mechanisms to reduce numbers and keep detection rates low and independent from the actual image size.

As final conclusion we state that a sophisticated selection of positions for embedding is not necessarily inferior to random selection.

Acknowledgement

The work on this paper was supported by the Air Force Office of Scientific Research under the research grant number FA8655-03-1-3A46. The U. S. Government is authorised to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Air Force Office of Scientific Research, or the U. S. Government.

References

1. Anderson, R. J., Petitcolas, F. A. P.: On the Limits of Steganography. IEEE Journal on Selected Areas in Communications **16** (1998) 474–481
2. Cachin, C.: An Information-Theoretic Model for Steganography. In: Aucsmith, D. (ed.): Information Hiding. Second International Workshop, LNCS 1525, Springer-Verlag, Berlin Heidelberg (1998) 306–318
3. Franz, E.: Steganography Preserving Statistical Properties. In: Petitcolas, F. A. P. (ed.): Information Hiding. 5th International Workshop, LNCS 2578, Springer-Verlag, Berlin Heidelberg (2003) 278–294

4. Franz, E.: personal communication (2003, Dec)
5. Franz, E., Pandit, V., Schneidewind, A.: Realization of Steganographic Algorithms and Discussion of Possible Problems. Unpublished working paper, Technische Universität Dresden (2004)
6. Fridrich, J., Goljan, M., Du, R.: Steganalysis Based on JPEG Compatibility. In: Tescher A. G., Vasudev B., Bove V. M., Jr. (eds.): Proceedings of SPIE, Multimedia Systems and Applications IV, Vol. 4518, Denver, CO, (2001) 275–280
7. Fridrich, J., Goljan, M., Du, R.: Reliable Detection of LSB Based Image Steganography. Proceedings of the ACM Workshop on Multimedia and Security (2001) 27–30
8. Ihaka, R., Gentlemen, R.: R—A Language for Data Analysis and Graphics. Journal of Computational Graphics and Statistics **5** (1996) 299–314
9. Johnson, N. F., Jajodia, S.: Steganalysis of Images Created Using Current Steganography Software. In: Aucsmith D. (ed.): Information Hiding. Second International Workshop, LNCS 1525, Springer-Verlag, Berlin Heidelberg (1998) 273–289
10. Katzenbeisser, S., Petitcolas, F. A. P.: On Defining Security in Steganographic Systems. Proceeding of SPIE, Security and Watermarking of Multimedia Contents IV, Vol. 4675. San Jose, California (2002) 50–56
11. Koops, B.-J.: Crypto Law Survey. Version 21.0, http://rechten.kub.nl/koops/cryptolaw/ (2002, Oct)
12. R Language for Statistical Computing, http://www.r-project.org
13. Sallee, P.: Model-Based Steganography. In: T. Kalker et al. (eds.): International Workshop on Digital Watermarking, LNCS 2939, Springer-Verlag, Berlin Heidelberg (2004) 154–167
14. Shannon, C. E.: A Mathematical Theory of Communication. Bell System Technical Journal **27** (1948) 379–423 623–656
15. Westfeld, A., Pfitzmann, A.: Attacks on Steganographic Systems. In: Pfitzmann, A. (ed.): Information Hiding. Third International Workshop, LNCS 1768, Springer-Verlag, Berlin Heidelberg (2000) 61–76
16. Zöllner, J., Federrath, H., Klimant, H., Pfitzmann, A., Piotraschke, R., Westfeld, A., Wicke, G., Wolf, G.: Modeling the Security of Steganographic Systems. In: Aucsmith D. (ed.): Information Hiding. Second International Workshop, LNCS 1525, Springer-Verlag, Berlin Heidelberg (1998) 344–354