The Square Root Law of Steganographic Capacity

Andrew D. Ker
Oxford University Computing Laboratory
Parks Road
Oxford OX1 3QD, UK
adk@comlab.ox.ac.uk

Tomáš Pevný
Dept. of Computer Science
SUNY Binghamton University
NY 13902-6000, USA
pevnak@gmail.com

Jan Kodovský
Dept. of Electrical and Computer Engineering
SUNY Binghamton University
NY 13902-6000, USA
jan.kodovsky@binghamton.edu

Jessica Fridrich
Dept. of Electrical and Computer Engineering
SUNY Binghamton University
NY 13902-6000, USA
fridrich@binghamton.edu

ABSTRACT

There are a number of recent information theoretic results demonstrating (under certain conditions) a sublinear relationship between the number of cover objects and their total steganographic capacity. In this paper we explain how these results may be adapted to the steganographic capacity of a single cover object, which under the right conditions should be proportional to the *square root* of the cover size. Then we perform some experiments using three genuine steganography methods in digital images, covering both spatial and DCT domains. Measuring detectability under four different steganalysis methods, for a variety of payload and cover sizes, we observe close accordance with a square root law.

Categories and Subject Descriptors

D.2.11 [**Software Engineering**]: Software Architectures—*information hiding*

General Terms

Security, Algorithms

Keywords

Steganographic Capacity, Benchmarking, Steganalysis, Steganography

1. INTRODUCTION

Clearly, the size of a cover object is a major factor in its capacity for hidden information. It is common for steganography and steganalysis literature to report data hiding as a *rate* (so many bits per second, bits per pixel, or, in transform domains, bits per useable coefficient) but it has been observed that rates are not comparable, and data hidden in a small cover is often less detectable than data hidden at the same rate in a large cover. Apart from making comparability of different authors' benchmarks difficult, it poses a fundamental question: how does secure steganographic capacity depend on the cover size, if this dependence is not proportional?

Some related theoretical work [13, 16] on the problem of *batch steganography* [12] (data hiding in multiple covers) demonstrates that, under certain conditions, the steganographic capacity of a batch of N cover objects is proportional only to $\sqrt{N}$. It suggests a square root law for steganographic capacity: a result in sharp contrast to capacity in noisy channels, which demonstrates that the theory of hidden information is rather different to the theory of information. In this paper we investigate that law, testing contemporary steganography and steganalysis methods to show that it appears valid in practice. The experiments must be designed carefully, and the law interpreted cautiously, because along with size there are other properties of cover images which significantly affect the detectability of payload.

The paper is structured as follows. In Sect. 2 we summarise some recent results about a square root law for steganographic capacity in multiple covers, and show that under (rather strong) conditions, they could apply to single covers too. In Sect. 3 we perform some experiments on spatial-domain steganography, examining how the accuracy of some leading steganalysis methods depends on the cover and payload size, and demonstrating close accordance with a square root law. In Sect. 4 we do the same for F5 steganography in JPEG images. Finally, in Sect. 5 we draw conclusions.

2. THE SQUARE ROOT LAW

Steganographic capacity is a loosely-defined concept, indicating the size of payload which may securely be embedded in a cover object using a particular embedding method. What constitutes "secure" embedding is a matter for debate, but we will argue that capacity should grow only as the square root of the cover size under a wide range of definitions of security.

Turning to the literature, even as long ago as 1996 we find references to the possibility of sublinear steganographic capacity. Anderson [1, §4.3] states:

> "Thanks to the Central Limit Theorem, the more covertext we give the warden, the better he will

MM&Sec'08, September 22–23, 2008, Oxford, UK.

> be able to estimate its statistics, and so the smaller the rate at which [the steganographer] will be able to tweak bits safely. The rate might even tend to zero..."

But it seems that this idea was not pursued and the rate of capacity was not quantified. Note, however, that the reference to the Central Limit Theorem already suggests that a square root relationship should be considered.

The first empirical results relating size of payload to size of cover can be found in [9, §2.1]: a fixed set of cover images was repeatedly rescaled to different sizes, fixed-rate payloads embedded, and one simple detector's performance measured. It was observed that the payload was more detectable in larger images, and concluded:

> "This is not contrary to the instinctively obvious fact that larger images can carry larger messages securely, but it does indicate that the increase is not proportional."

Further investigation in [2, §4.1] formulated statisical tests to distinguish between different possible relationships between cover size and detector dispersion (the inverse of which is a proxy for detectability). Of those tested, the best fitting was a log-proportional model for dispersion, and a square-root model the second best fit. The latter would suggest a square root law for capacity, the former something asymptotically greatly.

These results should not be taken at face value because the differently-sized covers were obtained by resampling. It is now known [3] that there are other highly significant factors which affect steganographic capacity, including the local variance of the cover. When an image undergoes significant downsampling, its local variance is usually increased, so the empirical results in [9, 2] had inadvertently caused, and did not control for, a confounding factor. Nonetheless, the results are strongly suggestive of sublinear capacity.

Notwithstanding these experimental results, there is theoretical work to show that information could in principle be hidden at a *linear* rate: in [24], codes are constructed which guarantee that the distribution of stego texts is identical to that of cover texts, and which do convey hidden messages proportional to the cover size. Indeed, it is entirely reasonable to believe that a cover source has a fixed entropy rate, in which case selecting a sequence of unaltered covers itself conveys information at a linear rate. However, such results depend on the steganographer knowing everything about their cover source. In the practice of steganography in digital media objects, a model of the source (if used at all) can only be an approximation, and there is always the possibility that a detector has a better model.

In a world where the steganography is not perfect, then, what is the relationship between cover size and capacity?

To our knowledge, the question is not addressed directly in the literature. However, a closely-related question has been studied recently. Instead of considering a single cover and its capacity in relation to its size, imagine a set of N cover objects amongst which a payload is to be spread. This is the problem of *batch steganography* introduced in [12]. In the batch setting, the capacity question becomes the relationship between the number of covers and their total secure capacity. In [12] we also find the first explicit conjecture that steganographic capacity follows a square root law.

Defining security by Neyman-Pearson style bounds on the performance of a *pooled* detector, which takes evidence from steganalysis of each individual object to decide whether a batch of N contains any payload, [13] proves a result about capacity. Under certain conditions, capacity is proportional to $\sqrt{N}$ in the following sense:

THEOREM 1. *Suppose that a detector maps objects to real scalars, and that the effect of embedding payload is cause a linear shift in the detector response; suppose also that the detector response density has infinite support, is at least twice continuously differentiable, and the second derivative of its logarithm is bounded below.*

If a steganographer embeds a total payload of M bits into N uniform cover objects, then

(1) If $M/\sqrt{N} \to \infty$ as $N \to \infty$ then there is a pooled detector which, for sufficiently large N, comes arbitrarily close to perfect detection.

(2) If $M/\sqrt{N} \to 0$ as $N \to \infty$ then, by spreading the payload equally between N covers, the performance of any pooled detector must become arbitrarily close to random for sufficiently large N.

(This paraphrases the result in [13].)

These conditions are certainly strong, particularly in assuming that embedded payload causes a linear shift in detector response (such an assumption is motivated by payload-size estimators in [12]). In [13] it is conjectured that this need hold only locally for payloads near zero. There is also an implicit assumption that the detector output is i.i.d. for covers, forcing a kind of uniformity on the cover objects themselves.

Although the hypotheses are strong, this was the first result to prove a square root capacity relationship between secure payload and cover size.

Another square root law for batch steganography is also proved, under different conditions, in [16]. There we have

THEOREM 2. *Suppose that the covers are N independent objects, that the KL divergence [18] between cover and stego objects with payload p is proportional to p^2, and that security is defined in terms of a bound on the total KL divergence between the sequence of covers and corresponding stego objects. Then the maximum secure payload M is $O(\sqrt{N})$.*

(This paraphrases Theorems 4 and 5 of [16].)

This version of the batch square root law has different hypotheses: it allows nonuniformity of the covers (as long as their characteristics are bounded in a suitable sense, see [16] for details) and transfers conditions on the detector into a hypothesis about KL divergence between cover and stego objects. The KL divergence assumption remains quite strong, although it can be justified by regularity conditions similar to those in [14], and it still requires independence of the cover objects.

How do these capacity results, for batch steganography, relate to the capacity of individual covers? Embedding in a single cover object can be modelled as an instance of the batch problem if we consider the cover to consist of a sequence of small regions, with the embedder having freedom to split the payload amongst the regions. However, the preceding theorems required independence of the component objects: not implausible for different cover objects, but un-

likely for regions within a single object because of image-wide effects of the cover source. Nonetheless, we might expect to see approximate independence of different regions, or at least believe that any dependencies are not exploited in steganalysis.

In practice, many steganalysis methods use only local measurements (based on 8×8 DCT blocks, or properties of small groups of adjacent pixels) and their aggregate operation treats the blocks as independent. If a steganalysis method can be expressed as some function of independent regions (whether or not it is expressly written in such a form) then it is an example of a pooled detector and obeys the batch capacity laws. This applies particularly to JPEG steganalysis, which tends to use statistics of 8×8 blocks without considering inter-block dependency.

In the absence of perfect steganography these discussions suggest that, all other things being equal, the secure steganographic capacity of a cover object should be proportional to the square root of the number of available embedding locations. Of course, the preceding discussion is certainly no proof of such a square root law in general. Indeed, we do not believe that a square root law will be a single theorem. Instead it is likely to constitute of suite of results, similar in style to Theorem 1, for a range of mathematical models of cover objects and embedding methods. For example, the batch steganography results deal with some cases where the cover objects are modelled as a sequence of independent random variables. That the square root law holds in general is a falsifiable, but probably unprovable, thesis.

Before testing this hypothesis empirically, we return briefly to the definition of capacity. It is not quite correct to speak of capacity as a bound on the size of payload because it is not payload itself which is detected by steganalysis. It is the changes induced by embedding which are detected, and capacity is more properly given by a bound on permissable changes; in simple embedding schemes where the changes are of fixed magnitude, it is the number of changes we should measure. This difference is important because of the existence of adaptive source codes [6], which can exploit freedom of choice of embedding locations to reduce the number of changes required. We will return to this question in Sect. 5, until then proceeding under the implicit assumption that embedding changes and payload size remain in fixed proportion.

3. EXPERIMENTAL INVESTIGATION: SPATIAL-DOMAIN STEGANOGRAPHY

We now conduct experiments to validate the square root law, concentrating on digital images. Payloads will be embedded, using a number of different embedding methods and various payload lengths, into cover images of different sizes. Then we will measure the ability of some recent steganalysis methods to detect the payload, and look for a square root relationship. In this section we will focus on spatial-domain embedding and detection methods; in the next we will consider DCT-domain steganography for which there are some additional challenges.

There are two difficulties to overcome in testing the theoretical result of Sect. 2. First, the caveat that capacity is a square root law *all other things being equal*. Other literature on the benchmarking of steganalysis [2, 3] has shown that there are cover properties other than size – local variance, saturation, prior image processing operations – which significantly affect the detectability of payload, and it is not possible to control or even determine them all. Therefore we cannot use sets of differently-sized covers from different sources to estimate how capacity depends on size: variations in the other properties may invalidate the results. Neither can we generate small cover images by downsampling large ones, because downsampled images have a higher semantic density so, usually, higher local variance. The solution is to use a single set of large covers and repeatedly crop down to smaller images. In an attempt to preserve other image characteristics, the cropped region can be chosen so that the average local variance (here measured by average absolute difference between neighbouring pixels) is as close as possible to that of the whole image. Our image libraries are not large enough to partition them into disjoint sets for cropping to different sizes, so we may observe correlation between the content of the different-sized cropped images, but this is not expected to cause significant effects in the experiments.

The second difficulty is to define "capacity". We can set a level of detection risk which the steganographer is prepared to accept, but (even apart from the fact that the level itself will be arbitrary) how to measure detectability? As discussed in [14] and [15], there are many different detection metrics found in the literature. For these experiments we will consider three metrics, two standard and one very recent:

(a) The area under the ROC curve of a binary classifier for the presence or absence of payload (AUR), unnormalized so that AUR = 0.5 corresponds to a random detector and AUR = 1 to perfect detection;

(b) The minimum sum of false positive and false negative errors for a binary classifier $P_E = \frac{1}{2}\min(f_p + f_n)$ (for comparability with other measures, $1 - P_E$ is used);

(c) Directly from the observed cover and stego distributions of steganalysis features, a recently-developed measure called Maximum Mean Discrepancy (MMD). Its key features are described in the Appendix.

In each case, higher values denote lower security.

Our first series of experiments was performed on never-compressed cover images. A set of 3000 images was downloaded from the NRCS website [19]: apparently scanned from film in full colour, these images vary slightly in size around approximately 2100×1500 pixels. We downsampled the images to a larger side of 1024 pixels, and reduced them to grayscale: the same set of images has been used by a number of steganalysis researchers. Nine sets each of 3000 grayscale cover images were then created by repeated cropping, selecting the crop region best to match the local variance of the original, to sizes 100×75, 200×150, ..., 900×675.

Random payload was embedded using simple LSB replacement (for payload smaller than maximum a random selection of embedding locations was used). We selected three different strategies for choosing the payload size according to cover size: embedding a fixed-size payload in all cover sets, embedding payload proportional to the square root of the number of cover pixels, and embedding payload proportional to the number of cover pixels. For each option, three different constants of proportionality were tested.

The method in [17] gives the currently-known best steganalysis of LSB replacement in never-compressed images,

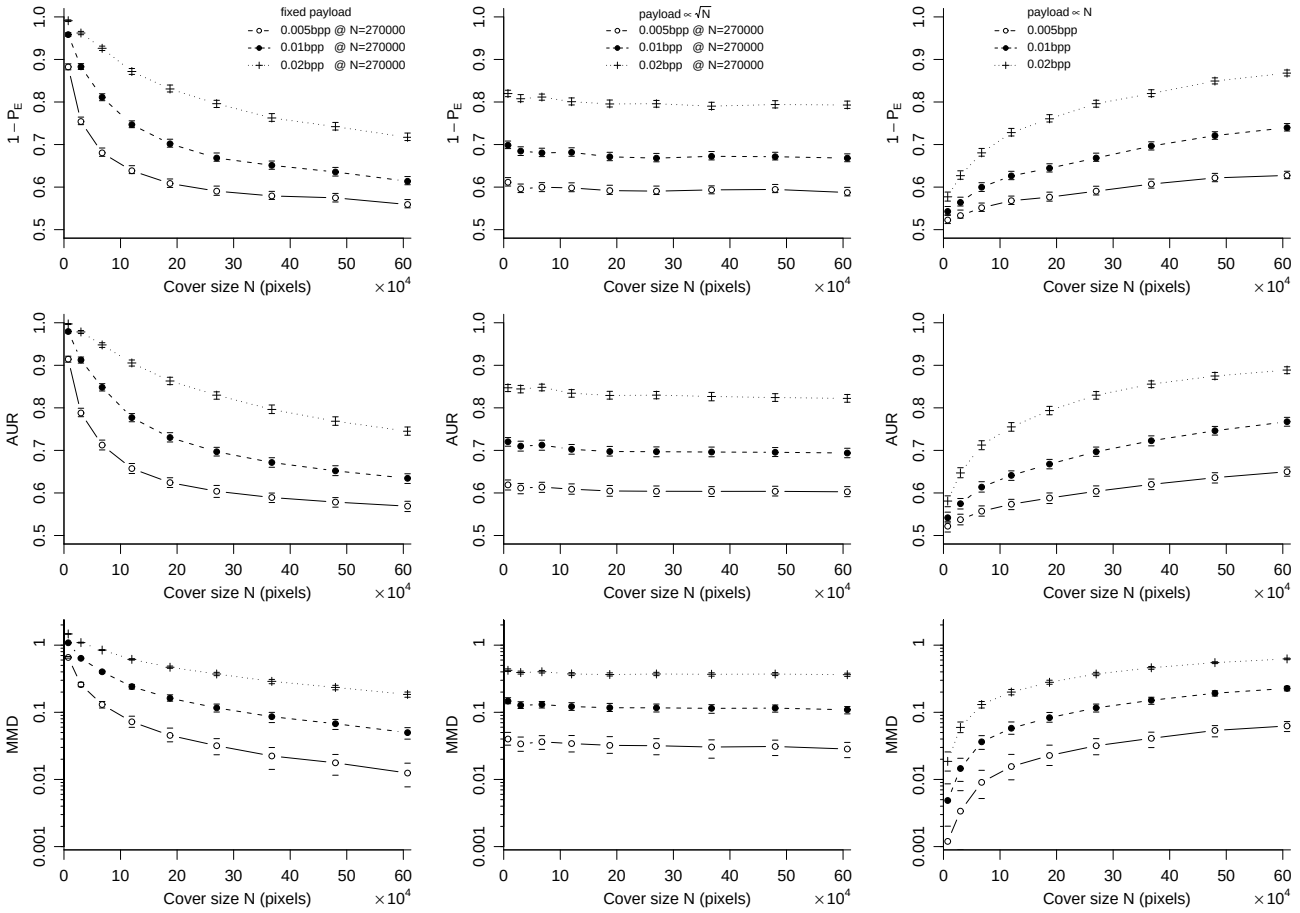

Figure 1: Detectability (y-axis, measured by AUR, $1 - P_E$, and MMD on a log scale) as a function of cover size N (x-axis) and payload size. 90% bootstrapped confidence intervals are indicated. Left, fixed payload size. Middle, payload proportional to $\sqrt{N}$. Right, proportional to N. LSB replacement steganography in never-compressed cover images, detected by method of [17].

and we applied it to each set of covers and stego images. The accuracies of the resulting detector for payload, as measured by AUR, P_E, and MMD, are displayed in Fig. 1, along with 90 % confidence intervals obtained using a simple resampling bootstrap. These experiments are in line with the theoretical predictions: whichever detectability metric is used, fixed-length payload becomes harder to detect in larger covers, fixed-proportion payload becomes easier to detect, and payload proportional to square root of cover size is (approximately) of constant detectability. At least these results suggest that square root capacity is much more plausible than proportionate capacity.

We repeated the same experiments with a set of 1600 images taken by the first author using a Minolta DiMAGE A1 camera in raw format at a resolution of 2000×1500, subsequently converted to grayscale and subject to JPEG compression (quality factor 80). The images were cropped to 16 different sizes between 100×75 and full size, again selecting the crop region to match the average local variance of the original. When cover images have been previously compressed, different detectors for LSB replacement have better performance than that in [17], so we used the *Triples* detector of [10].

Charts analogous to those in Fig. 1, for the compressed cover images and Triples steganalysis, are displayed in Fig. 2 and we can draw similar conclusions: secure payload is certainly not constant, nor proportional to cover size, but appears to be approximately proportional to the square root of the cover size. More visible in this second set of experiments are artefacts in the charts for very small cover sizes, but these are to be expected if the theoretical results are only asymptotic for large covers.

Finally, we tested an alternative method of spatial-domain LSB embedding known as LSB matching, or ± 1 embedding. It does not have the structural flaws of LSB replacement, and seems much more difficult to detect. For the detector, we used the method known as the *adjacency HCF COM* found in [11], but this detector is still quite weak: payloads as small as those in the previous two experiments are undetectable, so we had to increase the payload sizes considerably. As a result, it was not possible to fit the payloads into very small covers (one cannot embed more than 1 bit per pixel using LSBs). We used the same 3000 never-compressed scanned images as for the first experiment, cropped down to ten sizes between 360×270 and 900×675. The resulting charts are displayed in Fig. 3.

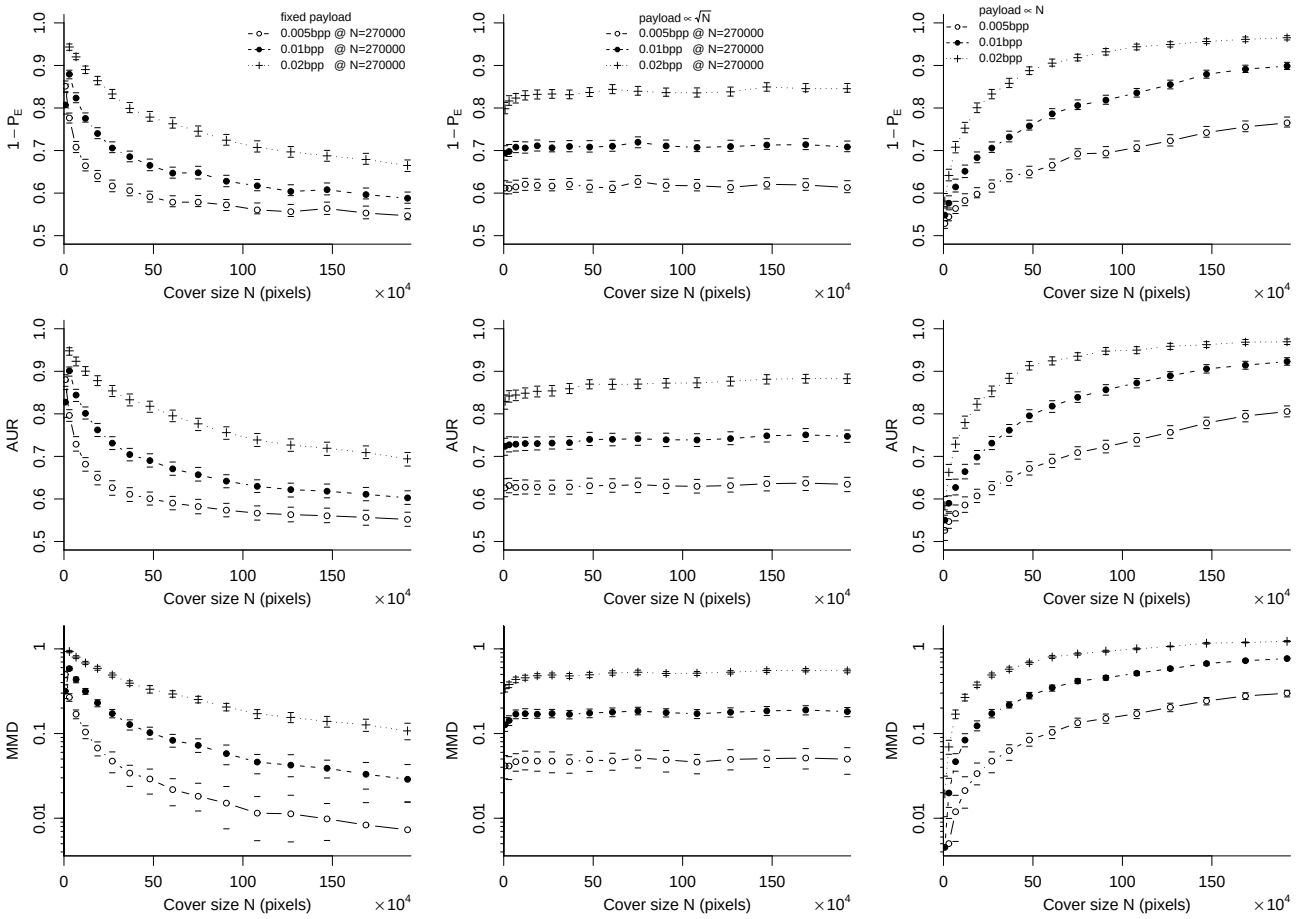

Figure 2: Detectability as a function of cover size, cf. Fig. 1. LSB replacement steganography in previously JPEG-compressed digital camera images, detected by method of [10].

Observe that the detector's performance remains very low: AUR is not much about 0.5 (which corresponds to a random detector), P_E not much below 0.5 (similarly), and MMD is near to zero (corresponding to identical distributions of cover and stego features) and, because we are digging in the detector noise, the bootstrap confidence intervals are wider. However, similar features are still apparent: falling detectability in larger covers when the payload is fixed and rising detectability when the payload is proportional to cover size. When the payload is proportional to the square root of the cover size, the detection metrics are *approximately* constant, although there is a suggestion that the detectability may be gradually decreasing.

To investigate more precisely how capacity depends on cover size we performed additional experiments: fixing on just one detection metric we set a bound on the risk to the steganographer (a minimum value of P_E) and determined the largest payload for which the detection bound can be met. This was accomplished by embedding 100 different payload sizes in each of the cover sets, measuring P_E for each combination and using linear interpolation to estimate P_E for intermediate payloads. Denoting cover size (pixels) by N and capacity (payload bits) by M, we can plot M against N on a log-log scale: if there is a relationship of the form $M \propto N^e$ then the points should fall in a straight line with slope e.

Figure 4 displays the results for each of the three detectors and cover sets in our experiments, with two different thresholds for P_E (in the case of LSB matching, we must set a very high threshold for P_E because the detector is so weak). In each case a straight line fit is determined by simple linear regression. When capacity is measured in this way, it does indeed appear to follow a relationship $M \propto N^e$, with values of e very close to 0.5. Even the line corresponding to $P_E = 0.45$ with the LSB matching detector would have slope close to 0.5 if the data points from the smallest image sets were discounted. Unfortunately we cannot use the standard least-squares tests for whether e differs *significantly* from 0.5, because the data points are not independent (they arise from images with overlapping content).

4. EXPERIMENTAL INVESTIGATION: JPEG STEGANOGRAPHY

We repeated the experiments of the previous section for steganography and steganalysis in JPEG images, to see whether the square root law still holds. A leading JPEG embedding method is F5 [25], and we used the improved version called no-shrinkage F5 (nsF5) [5], which has the same embedding operation but uses wet paper codes [4] to remove a

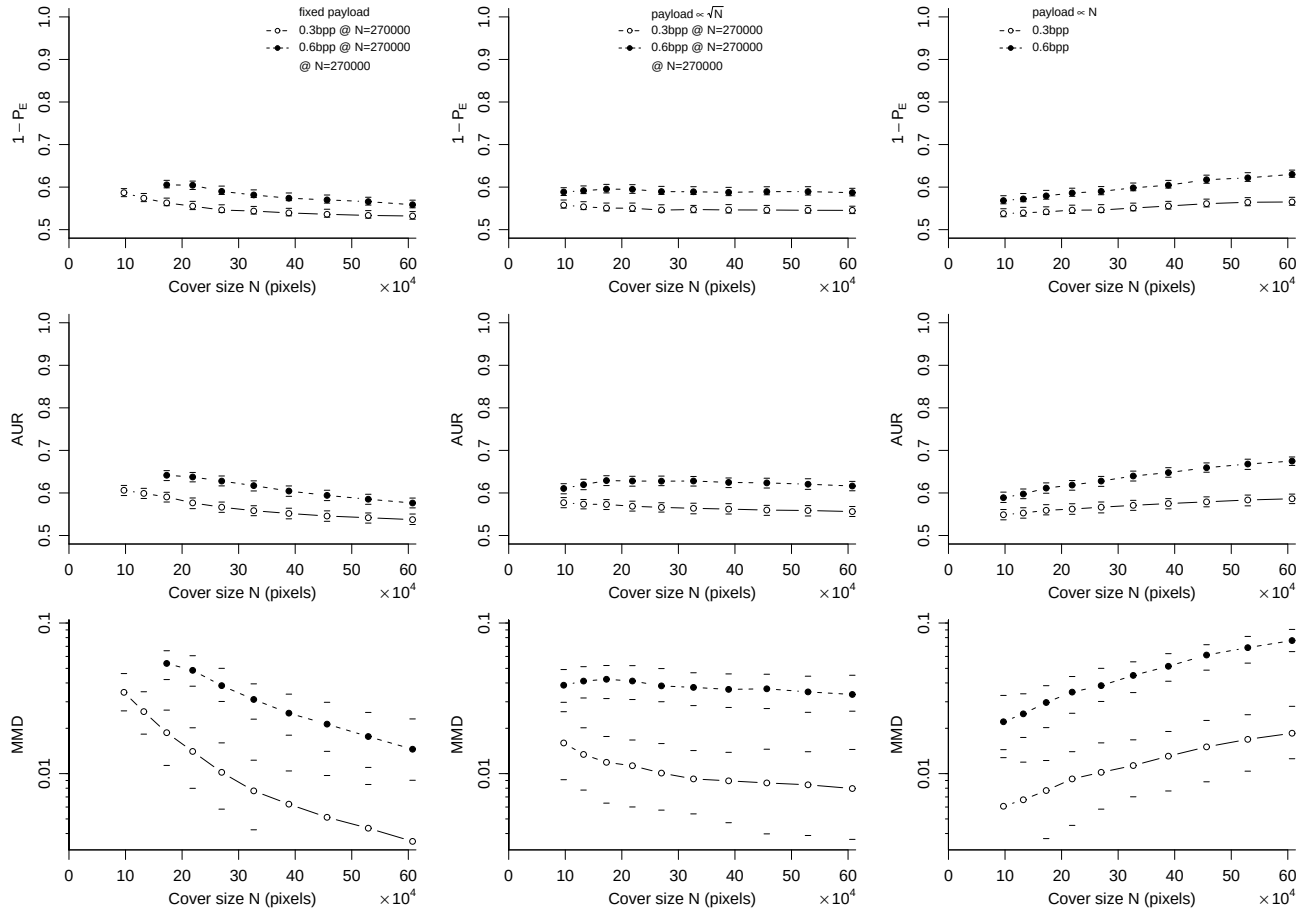

Figure 3: Detectability as a function of cover size, cf. Fig. 1. LSB matching steganography in never-compressed scanned images, detected by method of [11].

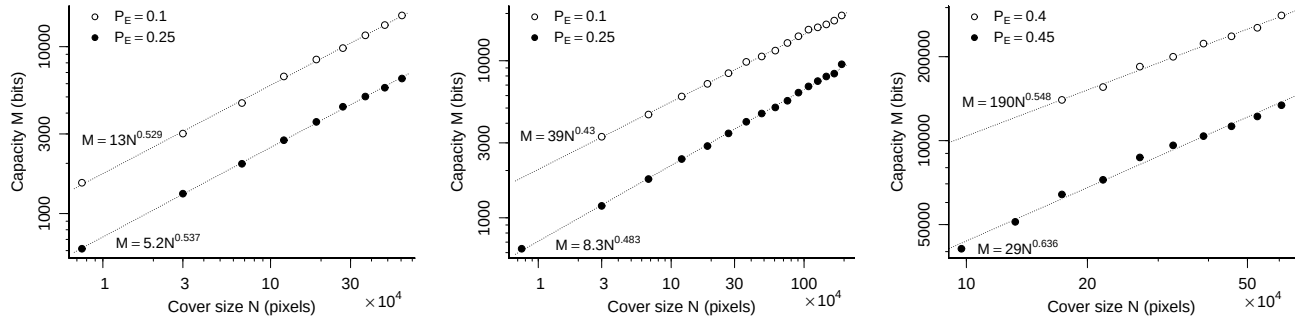

Figure 4: Capacity (y-axes, determined by limit on P_E) as a function of cover size (x-axis), log-log scale, with best-fit trend lines. Three different steganography/steganalysis methods displayed. Left, LSB replacement in never-compressed images detected by [17]; middle, LSB replacement in previously JPEG-compressed images detected by [10]; right, LSB matching detected by [11].

statistical anomaly where the absolute value of DCT coefficients tended to be reduced. F5 and nsF5 have an optional matrix embedding [6] feature, which was disabled because it introduces non-linearity between the payload and the number of embedding changes [5].

Measuring the *size* of a JPEG image is not as simple as counting pixels. After lossy compression, many of the DCT coefficients become zero and do not convey content: these coefficients cannot be used for embedding. Therefore we define the steganographic size as the total number of nonzero DCT coefficients (abbreviated nc). This is a generally-accepted measure, although some authors also discount DC coefficients.

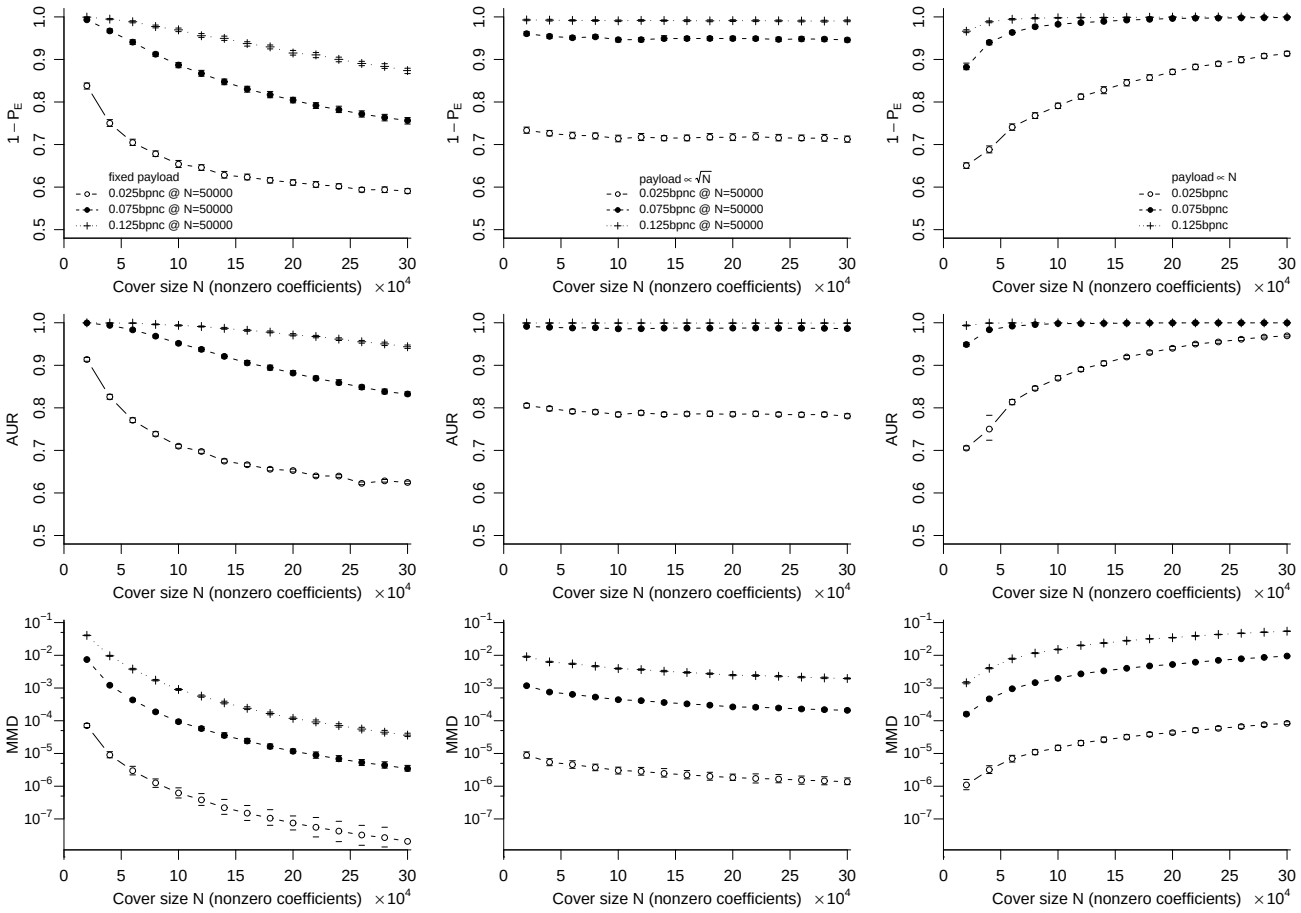

Figure 5: Detectability as a function of cover size (nonzero DCT coefficients). No-shrinkage F5 steganography with matrix embedding disabled, in JPEG covers, detected by method of [21].

We began with approximately 9200 never-compressed images of different sizes, and from them cropped 15 sets of cover images each with a specified number of nonzero coefficients, $2 \cdot 10^4$, $4 \cdot 10^4, \ldots,$ $30 \cdot 10^4$, all under JPEG compression with quality factor 80. None of the images were double-compressed. As in the spatial-domain experiments, cropping was favored over scaling: the latter produces images with a higher number of nonzero DCT coefficients on higher frequencies, so statistics of DCT coefficients in scaled images vary substantially with cover size. Also paralleling the experiments in the previous section, we chose the crop region to preserve some other characteristics of the cover. In the case of JPEG images, we attempted to preserve the proportion of nonzero DCT coefficients.

In each set of covers, a random message was embedded using the nsF5 algorithm. As before, our strategies for choosing the payload were to embed a fixed size payload into all cover sets, to embed payload proportional to the square root of the number of nonzero coefficients, and to embed payload proportionally to the number of nonzero coefficients.

The combination of Support Vector Machine (SVM) classifiers [22] with a Gaussian kernel and so-called *merged feature set* [21] is the state of art general purpose steganalytic system for JPEG images. We measured detectability using SVMs trained specifically to each combination of cover and payload size: for each such combination, 6000 images were selected at random from the available set of 9200, split into disjoint sets of 3500 for training and 2500 for testing. In the training stage, the 3500 cover images and 3500 corresponding stego images were used; similarly in the testing stage, the 2500 cover images and 2500 corresponding stego images were all classified by the SVM. The training and testing of the SVM classifiers was repeated 100 times with different random selections of training and testing sets, and the overall AUR and $1 - P_E$ metrics computed for the resulting binary classifiers.

Additionally, the MMD between the "merged feature set" vectors in cover and stego images was computed. Again, 6000 images were selected at random, this time partitioned into disjoint sets of 3000 covers and 3000 stego images (disjoint sets are necessary for good MMD estimation, see Appendix). This was repeated 100 times with random allocations of cover and stego images: increasing the accuracy of the estimate, and also allowing us to estimate rough bootstrap confidence intervals. Prior to computing MMD, the vectors were normalized so that each cover feature had zero mean and unit variance: note that, although the MMD kernel γ parameter (see Appendix) is fixed for all cover sizes, the normalization parameters are determined separately for each set. This proved necessary because we observed great

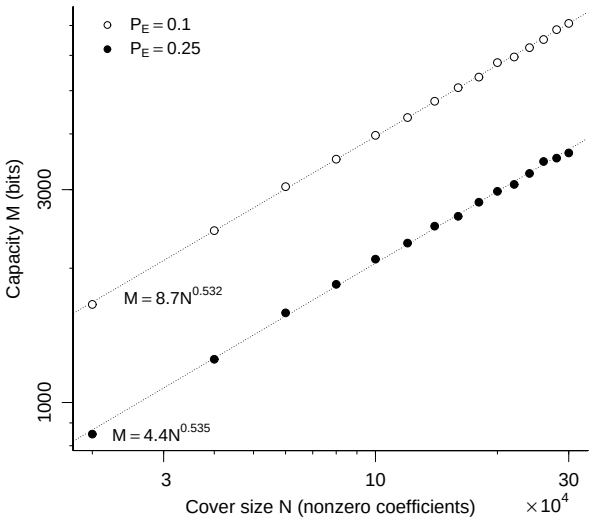

Figure 6: Capacity (y-axes, determined by limit on P_E) as a function of cover size (x-axis), log-log scale, with best-fit trend lines. No-shrinkage F5 in JPEG covers.

variability in the raw feature distributions, as the cover size varied.

The results of the experiment (Figure 5) confirm the theoretical predictions, and are similar to the results presented in Section 3. For fixed (respectively, linear) payload, by any metric the detectability increases (resp. decreases) with the cover size, and for payload proportional to the square root of nc the detectability is approximately constant, albeit with a barely-visible downwards trend. It is not known why the MMD measure shows this as a slightly stronger effect than AUR or P_E.

Following Sect. 3, the next experiment was to find payload such that the probability of error P_E matches a certain level. The search for the payload was carried under the reasonable assumption that the detectability increases with the payload size. The P_E measure at each given payload was estimated by the accuracy of the classifier (again, a SVM with a Gaussian kernel employing "merged" features) targeted to a given combination of nc and payload. The training and testing conditions were the same as in the previous experiment. Even though repeated training of the classifier is very time consuming, this approach was favoured because it provides good estimates of P_E.

Figure 6 shows maximum payload M plotted against nc N in log-log scale for $P_E = 0.1$ and $P_E = 0.25$. Payloads were identified within 1% accuracy of the desired P_E level. In both cases, the graph shows a close accordance with a straight line and the slope of the line is close to 0.5. This shows that the capacity of JPEG images for nsF5 (without matrix embedding) grows with square root of the number of nonzero DCT coefficients.

5. CONCLUSIONS

In this work we have surveyed the literature relating steganographic capacity to cover size, and argued that the square root law proved for batch steganography may also apply to the case of individual covers. There are suggestions, from the literature on random processes, that the square root law should also hold in rather general circumstances for Markov chains: this would be powerful additional evidence for square root capacity in general, and is the subject of future research.

Using carefully-designed experiments, which as far as possible isolate the effect of cover size from other cover properties, we tested the square root law for a number of steganography schemes, using contemporary steganalysis detectors. Close adherence to the law was observed.

It is not widely known that the secure capacity of a cover is proportional only to the square root of its size (where size should be measured by available embedding locations), in the absence of perfect steganography. It seems to be of fundamental importance to the practice of steganography, and could be particularly vital for the design of steganographic file systems, where the user might expect to be given an indication of secure capacity.

However, when interpreting the square root law we must take care not to ignore other important factors which contribute to capacity. In practice, properties of cover images such as saturation, local variance, and prior JPEG compression or image processing operations have been shown to have significant effects on detectability of payload [2, 3]. We cannot simply conclude that, because one cover is twice as larger as another, it can carry $\sqrt{2}$ times the payload at an equivalent risk. The law applies *other all things being equal* and, as the difficulties constructing suitable experiments to test the law illustrate, rarely are cover images equal.

We also emphasise that the law truly applies not to raw payload size but to the embedding changes caused. In some embedding schemes these quantities are not proportional. For example, using syndrome coding [6] and binary embedding operations it is possible to design embedding codes for which the number of embedding changes c and payload size M approaches asymptotically the bound $c \geq NH^{-1}(M/N)$, where H is the binary entropy function. The consequence of an asymptotic limit $c = O(\sqrt{N})$ is then $M = O(\sqrt{N}\log N)$. A parallel result is found in [16]. It would appear that, fundamentally, steganographic payload capacity is of order $\sqrt{N}\log N$. This is a curious outcome.

One could argue that, because of the square root law, researchers should cease to report payloads measured in bits per pixel, bits per second, bits per nonzero coefficient, etc: the correct units should perhaps be bits per square root pixel and so on. However, such a change would still not allow comparability of different authors' benchmarks, because of the other factors affecting detectability; unless different authors use covers from the same source, their results cannot be exactly comparable in any case.

For future research, it may be valuable to repeat experiments analogous to those in this paper for yet more steganography and steganalysis methods, including in domains other than digital images. Ideally, experiments would be conducted using so large a library of cover objects that the subsets of objects of different size came from disjoint originals. Then it would be possible to perform statistical tests for whether the capacity exponent differs significantly from 0.5. Our investigations also revealed (unreported) interactions between cover size and feature statistics, whose cause is yet to be identified.

6. ACKNOWLEDGMENTS

The first author is a Royal Society University Research Fellow. Thanks are due to Rainer Böhme for advice and discussions.

The work of J. Fridrich, J. Kodovský, and T. Pevný was supported by Air Force Office of Scientific Research under the research grant number FA9550-08-1-0084. The U.S. Government is authorized 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 AFOSR or the U.S. Government.

7. REFERENCES

[1] R. Anderson. Stretching the limits of steganography. In *Proc. 1st Information Hiding Workshop*, volume 1174 of *Springer LNCS*, pages 39–48, 1996.

[2] R. Böhme. Assessment of steganalytic methods using multiple regression models. In *Proc. 7th Information Hiding Workshop*, volume 3727 of *Springer LNCS*, pages 278–295, 2005.

[3] R. Böhme and A. Ker. A two-factor error model for quantitative steganalysis. In *Security, Steganography and Watermarking of Multimedia Contents VIII*, volume 6072 of *Proc. SPIE*, pages 59–74, 2006.

[4] J. Fridrich, M. Goljan, and D. Soukal. Wet paper codes with improved embedding efficiency. *IEEE Trans. Information Forensics and Security*, 1(1):102–110, 2006.

[5] J. Fridrich, T. Pevný, and J. Kodovský. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In *Proc. 9th ACM Workshop on Multimedia and Security*, pages 3–14, 2007.

[6] J. Fridrich and D. Soukal. Matrix embedding for large payloads. *IEEE Trans. Information Forensics and Security*, 1(3):390–394, 2006.

[7] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola. A kernel method for the two-sample-problem. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 513–520. MIT Press, Cambridge, MA, 2007.

[8] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola. A kernel method for the two-sample-problem. Technical report, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 2007. MPI Technical Report 157.

[9] A. Ker. Improved detection of LSB steganography in grayscale images. In *Proc. 6th Information Hiding Workshop*, volume 3200 of *Springer LNCS*, pages 97–115, 2004.

[10] A. Ker. A general framework for the structural steganalysis of LSB replacement. In *Proc. 7th Information Hiding Workshop*, volume 3727 of *Springer LNCS*, pages 296–311, 2005.

[11] A. Ker. Steganalysis of LSB matching in grayscale images. *IEEE Signal Processing Letters*, 12(6):441–444, 2005.

[12] A. Ker. Batch steganography and pooled steganalysis. In *Proc. 8th Information Hiding Workshop*, volume 4437 of *Springer LNCS*, pages 265–281, 2006.

[13] A. Ker. A capacity result for batch steganography. *IEEE Signal Processing Letters*, 14(8):525–528, 2007.

[14] A. Ker. The ultimate steganalysis benchmark? In *Proc. 9th ACM Workshop on Multimedia and Security*, pages 141–148, 2007.

[15] A. Ker. Benchmarking steganalysis. In C.-T. Li, editor, *Multimedia Forensics and Security*. IGI Global, 2008.

[16] A. Ker. Steganographic strategies for a square distortion function. In *Security, Forensics, Steganography and Watermarking of Multimedia Contents X*, volume 6819 of *Proc. SPIE*, 2008.

[17] A. Ker and R. Böhme. Revisiting weighted stego-image steganalysis. In *Security, Forensics, Steganography and Watermarking of Multimedia Contents X*, volume 6819 of *Proc. SPIE*, 2008.

[18] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[19] NRCS photo gallery. http://photogallery.nrcs.usda.gov/, accessed April 2004.

[20] T. Pevný and J. Fridrich. Benchmarking for steganography. To appear in *Proc. 10th Information Hiding Workshop*, 2008.

[21] T. Pevný and J. Fridrich. Merging Markov and DCT features for multi-class JPEG steganalysis. In *Security, Steganography and Watermarking of Multimedia Contents IX*, volume 6505 of *Proc. SPIE*, pages 3 1–3 14, 2007.

[22] J. Shawe-Taylor and N. Cristianini. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[23] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2:67–93, 2001.

[24] Y. Wang and P. Moulin. Perfectly secure steganography: Capacity, error exponents, and code constructions. *To appear in IEEE Trans. Information Theory*, 2008.

[25] A. Westfeld. F5 – a steganographic algorithm: High capacity despite better steganalysis. In *Proc. 4th Information Hiding Workshop*, volume 2137 of *Springer LNCS*, pages 289–302, 2001.

APPENDIX

A. THE MMD MEASURE

Maximum Mean Discrepancy (MMD) [7] is a recently-developed measure of difference between probability distributions. If X and Y are random variables with the same domain $\mathcal{X}$ then their MMD is defined as

$$\max \left| \mathbf{E}[f(X)] - \mathbf{E}[f(Y)] \right|, \tag{1}$$

where the maximum is taken over all mappings $f : \mathcal{X} \mapsto \mathbb{R}$ from a unit ball $\mathcal{F}$ in a Reproducing Kernel Hilbert Space (RKHS). Although not a true metric, the MMD is symmetric, nonnegative, and zero only when X and Y have the same distribution.

For technical reasons it is simpler to use the square of the MMD measure in (1) and in this paper we always report squared MMD values. Given n independent observations $(x_1, \ldots, x_n)$ of X and a further n independent observations $(y_1, \ldots, y_n)$ of Y, the (squared) MMD may be estimated by

$$\frac{1}{n(n-1)} \sum_{i \neq j} \sum k(x_i, x_j) + k(y_i, y_j) - 2k(x_i, y_j)$$

where k is a bounded universal kernel $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ that defines the dot product in the RKHS [23]. The variance of the estimator decreases as $1/\sqrt{n}$, almost independently of the dimension of the random variables [8], and can also be improved by bootstrapping. MMD has been used for comparing security of stego-schemes in [20].

In this paper we measure MMD with respect to a Gaussian kernel

$$k(x, y) = \exp(-\gamma \|x - y\|^2),$$

with the width parameter γ set to η^{-2}, where η is the median of the L_2-distances between (normalized) features in a pooled set of all cover images. This choice is justified in [20]. Note that, for direct comparison of MMD values obtained from experiments on different cover sets, the γ parameter should remain fixed.