Advanced Statistical Steganalysis

Bearbeitet von
Rainer Böhme

1. Auflage 2010. Buch. xvi, 288 S. Hardcover
ISBN 978 3 642 14312 0
Format (B x L): 15,5 x 23,5 cm
Gewicht: 619 g

Chapter 2
Principles of Modern Steganography and Steganalysis

The first work on digital steganography was published in 1983 by cryptographer Gustavus Simmons [217], who formulated the problem of steganographic communication in an illustrative example that is now known as the *prisoners' problem*[1]. Two prisoners want to cook up an escape plan together. They may communicate with each other, but all their communication is monitored by a warden. As soon as the warden gets to know about an escape plan, or any kind of scrambled communication in which he suspects one, he would put them into solitary confinement. Therefore, the inmates must find some way of hiding their secret messages in inconspicuous cover text.

2.1 Digital Steganography and Steganalysis

Although the general model for steganography is defined for arbitrary communication channels, only those where the cover media consist of multimedia objects, such as image, video or audio files, are of practical relevance.[2] This is so for three reasons: first, the cover object must be large compared to the size of the secret message. Even the best-known embedding methods do not allow us to embed more than 1% of the cover size securely (cf. [87, 91] in conjunction with Table A.2 in Appendix A). Second, indeterminacy[3] in the cover is necessary to achieve steganographic security. Large objects without indeterminacy, e.g., the mathematical constant π at very high precision, are unsuitable covers since the warden would be able to verify their regular

[1] The prisoners' problem should not be confused with the better-known prisoners' dilemma, a fundamental concept in game theory.

[2] Artificial channels and 'exotic' covers are briefly discussed in Sects. 2.6.1 and 2.6.5, respectively.

[3] Unless otherwise stated, *indeterminacy* is used with respect to the uninvolved observer (warden) throughout this book. The output of indeterministic functions may be deterministic for those who know a (secret) internal state.

structure and discover traces of embedding. Third, transmitting data that contains indeterminacy must be plausible. Image and audio files are so vital nowadays in communication environments that sending such data is inconspicuous.

As in modern cryptography, it is common to assume that Kerckhoffs' principle [135] is obeyed in digital steganography. The principle states that the steganographic algorithms to embed the secret message into and extract it from the cover should be public. Security is achieved solely through secret keys shared by the communication partners (in Simmons' anecdote: agreed upon before being locked up). However, the right interpretation of this principle for the case of steganography is not always easy, as the steganographer may have additional degrees of freedom [129]. For example, the selection of a cover has no direct counterpart in standard cryptographic systems.

2.1.1 Steganographic System

Figure 2.1 shows the baseline scenario for digital steganography following the terminology laid down in [193]. It depicts two parties, sender and recipient, both steganographers, who communicate covertly over the public channel. The sender executes function $\mathsf{Embed} : \mathcal{M} \times \mathcal{X}^* \times \mathcal{K} \to \mathcal{X}^*$ that requires as inputs the secret message $\boldsymbol{m} \in \mathcal{M}$, a plausible cover $\boldsymbol{x}^{(0)} \in \mathcal{X}^*$, and the secret key $\boldsymbol{k} \in \mathcal{K}$. $\mathcal{M}$ is the set of all possible messages, $\mathcal{X}^*$ is the set of covers transmittable over the public channel and $\mathcal{K}$ is the key space. Embed outputs a stego object $\boldsymbol{x}^{(m)} \in \mathcal{X}^*$ which is indistinguishable from (but most likely not identical to) the cover. The stego object is transmitted to the recipient who runs $\mathsf{Extract} : \mathcal{X}^* \times \mathcal{K} \to \mathcal{M}$, using the secret key $\boldsymbol{k}$, to retrieve the secret message $\boldsymbol{m}$. Note that the recipient does not need to know the original cover to extract the message. The relevant difference between covert and encrypted communication is that for covert communication it is hard or impossible to infer the *mere existence* of the secret message from the observation of the stego object without knowledge of the secret key.

The combination of embedding and extraction function for a particular type of cover, more formally the quintuple $(\mathcal{X}^*, \mathcal{M}, \mathcal{K}, \mathsf{Embed}, \mathsf{Extract})$, is called *steganographic system*, in short, *stego system*.[4]

[4] This definition differs from the one given in [253]: Zhang and Li model it as a sextuple with separate domains for covers and stego objects. We do not follow this definition because the domain of the stego objects is implicitly fixed for given sets of covers, messages and keys, and two transformation functions. Also, we deliberately exclude distribution assumptions for covers from our system definition.

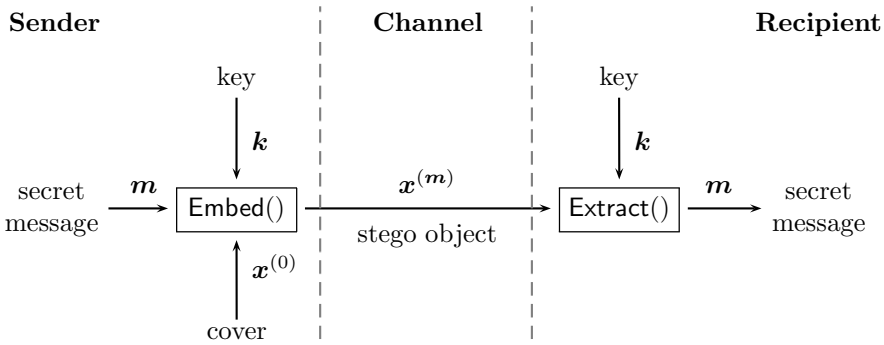


Fig. 2.1: Block diagram of baseline steganographic system

2.1.2 Steganalysis

The security of a steganographic system is defined by its strength to defeat detection. The effort to detect the presence of steganography is called *steganalysis*. The steganalyst (i.e., the warden in Simmons' anecdote) is assumed to control the transmission channel and watch out for suspicious material [114]. A steganalysis method is considered as successful, and the respective steganographic system as 'broken', if the steganalyst's decision problem can be solved with higher probability than random guessing [33].

Note that we have not yet made any assumptions on the computational complexity of the algorithms behind the functions of the steganographers, Embed and Extract, and the steganalyst's function $\mathsf{Detect} : \mathcal{X}^* \to \{\mathsf{cover}, \mathsf{stego}\}$. It is not uncommon that the steganalyst's problem can theoretically be solved with high probability; however, finding the solution requires vast resources. Without going into formal details, the implicit assumption for the above statements is that for an operable steganographic system, embedding and extraction are computationally easy whereas reliable detection requires considerably more resources.

2.1.3 Relevance in Social and Academic Contexts

The historic roots of steganography date back to the ancient world; the first books on the subject were published in the 17th century. Therefore, the art is believed to be older than cryptography. We do not repeat the phylogenesis of covert communication and refer to Kahn [115], Petitcolas et al. [185]

or, more comprehensively, Kipper [139, Chapter 3], who have collected numerous examples of covert communication in the pre-digital age. Advances in modern digital steganography are relevant for academic, engineering, national security and social reasons. For society at large, the existence of secure steganography is a strong argument for the opponents of crypto regulation, a debate that has been fought in Germany in the 1990s and that reappears on the agendas of various jurisdictions from time to time [63, 142, 143]. Moreover, steganographic mechanisms can be used in distributed peer-to-peer networks that allow their users to safely evade Internet censorship imposed by authoritarian states. But steganography is also a 'dual use' technique: it has applications in defence, more precisely in covert field communication and for hidden channels in cyber-warfare tools. So, supposedly intelligence agencies are primarily interested in steganalysis. Steganography in civilian engineering applications can help add new functionality to legacy protocols while maintaining compatibility (the security aspect is subordinated in this case) [167]. Some steganographic techniques are also applicable in digital rights management systems to protect intellectual property rights of media data. However, this is mainly the domain of digital watermarking [42], which is related to but adequately distinct from pure steganography to fall beyond the scope of this book. Both areas are usually subsumed under the term 'information hiding' [185].[5] Progress in steganography is beneficial from a broader academic perspective because it is closely connected to an ever better understanding of the stochastic processes behind cover data, i.e., digital representations of natural images and sound. Refined models, for whatever purpose, can serve as building blocks for better compression and recognition algorithms. Steganography is interdisciplinary and touches fields of computer security, particularly cryptography, signal processing, coding theory, and machine learning (pattern matching). Steganography is also closely conected (both methodologically but also by an overlapping academic community) to the emerging field of multimedia forensics. This branch develops [177] and challenges [98, 140] methods to detect forgeries in digital media.

2.2 Conventions

Throughout this book, we use the following notation. Capital letters are reserved for random variables X defined over the domain $\mathcal{X}$. Sets and multisets are denoted by calligraphic letters $\mathcal{X}$, or by double-lined capitals for special sets $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, etc. Scalars and realisations of random variables are printed in lower case, x. Vectors of n random variables are printed in boldface (e.g.,

[5] Information hiding as a subfield of information security should not be confused with information hiding as a principle in software engineering, where some authors use this term to describe techniques such as abstract data types, object orientation, and components. The idea is that lower-level data structures are hidden from higher-level interfaces [181].

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ takes its values from elements of the product set $\mathcal{X}^n$). Vectors and matrices, possibly realisations of higher-dimensional random variables, are denoted by lower-case letters printed in boldface, $\boldsymbol{x}$. Their elements are annotated with a subscript index, x_i for vectors and $x_{i,j}$ for matrices. Subscripts to boldface letters let us distinguish between realisations of a random vector; for instance, $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are two different secret messages. Functions are denoted by sequences of characters printed in sans serif font, preceded by a capital letter, for example, $\mathsf{F}(x)$ or $\mathsf{Embed}(\boldsymbol{m}, \boldsymbol{x}^{(0)}, \boldsymbol{k})$.

No rule without exception: we write $\boldsymbol{k}$ for the key, but reuse scalar k as an index variable without connection to any element of a vector of key symbols. Likewise, N is used as alternative constant for dimensions and sample sizes, not as a random variable. $\boldsymbol{I}$ is the identity matrix (a square matrix with 1s on the main diagonal and 0s elsewhere), not a random vector. Also $\mathcal{O}$ has a double meaning: as a set in sample pair analysis (SPA, Sect. 2.10.2), and elsewhere as the complexity-theoretic Landau symbol $\mathcal{O}(n)$ with denotation 'asymptotically bounded from above'.

We use the following conventions for special functions and operators:

- **Set theory** $\mathfrak{P}$ is the power set operator and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$.
- **Matrix algebra** The inverse of matrix $\boldsymbol{x}$ is $\boldsymbol{x}^{-1}$; its transposition is $\boldsymbol{x}^{\mathsf{T}}$. The notation $\mathbf{1}_{i \times j}$ defines a matrix of 1s with dimension i (rows) and j (columns). Operator $\otimes$ stands for the Kronecker matrix product or the outer vector product, depending on its arguments. Operator $\odot$ denotes element-wise multiplication of arrays with equal dimensions.
- **Information theory** $\mathsf{H}(X)$ is the Shannon entropy of a discrete random variable or empirical distribution (i.e., a histogram). $\mathsf{D}_{\mathsf{KL}}(X, Y)$ is the relative entropy (Kullback–Leibler divergence, KLD [146]) between two discrete random variables or empirical distributions, with the special case $\mathsf{D}_{\mathsf{bin}}(u, v)$ as the binary relative entropy of two distributions with parameters $(u, 1-u)$ and $(1-v, v)$. $\mathsf{D}_{\mathsf{H}}(\boldsymbol{x}, \boldsymbol{y})$ is the Hamming distance between two discrete sequences of equal length.
- **Probability calculus** $\mathsf{Prob}(x)$ denotes the probability of event x, and $\mathsf{Prob}(x|y)$ is the probability of x conditionally on y. Operator $\mathsf{E}(X)$ stands for the expected value of its argument X. $X \sim \mathcal{N}(\mu, \sigma)$ means that random variable X is drawn from a Gaussian distribution with mean μ and standard deviation σ. Analogously, we write $\boldsymbol{\mathcal{N}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the multivariate case with covariance matrix $\boldsymbol{\Sigma}$. When convenient, we also use probability spaces $(\Omega, \mathcal{P})$ on the right-hand side of operator '$\sim$', using the simplified notation $(\Omega, \mathcal{P}) = (\Omega, \mathfrak{P}(\Omega), \mathcal{P})$ since the set of events is implicit for countable sample spaces. We write the uniform distribution over the interval $[a, b]$ as $\mathcal{U}_a^b$ in the continuous case and as $\ddot{\mathcal{U}}_a^b$ in the discrete case (i.e., all integers $i : a \leq i \leq b$ are equally probable). Further, $\mathcal{B}(n, \pi)$ stands for a binomial distribution as the sum of n Bernoulli trials over $\{0, 1\}$ with probability to draw a 1 equal to π. Unless otherwise stated,

the hat annotation $\hat{x}$ refers to an estimate of a true parameter x that is only observable indirectly through realisations of random variables.

We further define a special notation for embedded content and write $\boldsymbol{x}^{(0)}$ for cover objects and $\boldsymbol{x}^{(1)}$ for stego objects. If the length of the embedded message is relevant, then the superscript may contain a scalar parameter in brackets, $\boldsymbol{x}^{(p)}$, with $0 \leq p \leq 1$, measuring the secret message length as a fraction of the total capacity of $\boldsymbol{x}$. Consistent with this convention, we write $\boldsymbol{x}^{(i)}$ if it is uncertain, but not irrelevant whether $\boldsymbol{x}$ represents a cover or a stego object. In this case we specify i further in the context. If we wish to distinguish the content of multiple embedded messages, then we write $\boldsymbol{x}^{(\boldsymbol{m}_1)}$ and $\boldsymbol{x}^{(\boldsymbol{m}_2)}$ for stego objects with embedded messages $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$, respectively. The same notation can also be applied to elements x_i of $\boldsymbol{x}$: $x_i^{(0)}$ is the ith symbol of the plain cover and $x_i^{(1)}$ denotes that the ith symbol contains a steganographic semantic. This means that this symbol is used to convey the secret message and can be interpreted by $\mathsf{Extract}$. In fact, $x_i^{(0)} = x_i^{(1)}$ if the steganographic meaning of the cover symbol already matches the respective part of the message. Note that there is not necessarily a one-to-one relation between message symbols and cover symbols carrying secret message information $x_i^{(1)}$, as groups of cover symbols can be interpreted jointly in certain stego systems (cf. Sect. 2.8.2).

Without loss of generality, we make the following assumptions in this book:

- The secret message $\boldsymbol{m} \in \mathcal{M} = \{0,1\}^*$ is a vector of bits with maximum entropy. (The Kleene closure operator * is here defined under the vector concatenation operation.) We assume that symbols from arbitrary discrete sources can be converted to such a vector using appropriate source coding. The length of the secret message is measured in bits and denoted as $|\boldsymbol{m}| \geq 0$ (as the absolute value interpretation of the $|x|$ operator can be ruled out for the message vector). All possible messages of a fixed length appear with equal probability. In practice, this can be ensured by encrypting the message before embedding.
- Cover and stego objects $\boldsymbol{x} = (x_1, \ldots, x_n)$ are treated as column vectors of integers, thus disregarding any 2D array structure of greyscale images, or colour plane information for colour images. So, we implicitly assume a homomorphic mapping between samples in their spatial location and their position in vector $\boldsymbol{x}$. Whenever the spatial relation of samples plays a role, we define specific mapping functions, e.g., $\mathsf{Right} : \mathbb{Z}^+ \to \mathbb{Z}^+$ between the indices of, say, a pixel x_i and its right neighbour x_j, with $j = \mathsf{Right}(i)$. To simplify the notation, we ignore boundary conditions when they are irrelevant.

2.3 Design Goals and Metrics

Steganographic systems can be measured by three basic criteria: capacity, security, and robustness. The three dimensions are not independent, but should rather be considered as competing goals, which can be balanced when designing a system. Although there is a wide consensus on the same basic criteria, the metrics by which they are measured are not unanimously defined. Therefore, in the following, each dimension is discussed together with its most commonly used metrics.

2.3.1 Capacity

Capacity is defined as the maximum length of a secret message. It can be specified in *absolute* terms (bits) for a given cover, or as *relative* to the number of bits required to store the resulting stego object. The capacity depends on the embedding function, and may also depend on properties of the cover $\boldsymbol{x}^{(0)}$. For example, least-significant-bit (LSB) replacement with one bit per pixel in an uncompressed eight-bit greyscale image achieves a net capacity of 12.5%, or slightly less if one takes into account that each image is stored with header information which is not available for embedding. Some authors would report this as 1 bpp (*bits per pixel*), where the information about the actual bit depths of each pixel has to be known from the context. Note that not all messages are maximum length, so bits per pixel is also used as a measure of capacity usage or *embedding rate*. In this work, we prefer the latter term and define a metric p (for 'proportion') for the length of the secret message relative to the maximum secret message length of a cover. Embedding rate p has no unit and is defined in the range $0 \leq p \leq 1$. Hence, for an embedding function which embeds one bit per cover symbol,

$$p = \frac{|\boldsymbol{m}|}{n} \quad \text{for covers} \quad \boldsymbol{x}^{(0)} \in \mathcal{X}^n. \tag{2.1}$$

However, finding meaningful measures for capacity and embedding rate is not always as easy as here. Some stego systems embed into compressed cover data, in which the achievable compression rate may vary due to embedding. In such cases it is very difficult to agree on the best denominator for the capacity calculation, because the size of the cover (e.g., in bytes, or in pixels for images) is not a good measure of the amount of information in a cover. Therefore, specific capacity measures for particular compression formats of cover data are needed. For example, F5, a steganographic algorithm for JPEG-compressed images, embeds by *decreasing* the file size almost monotonically with the amount of embedded bits [233]. Although counterintuitive at first sight, this works by reducing the image quality of the lossy compressed image

Table 2.1: Result states and error probabilities of a binary detector

	Reality	
Detector output	plain cover	stego object
plain cover	correct rejection $1-\alpha$	miss β
stego object	false positive α	correct detection $1-\beta$

further below the level of distortion that would occur without steganographic content. As a result, bpc (*bits per nonzero DCT coefficient*) has been proposed as a capacity metric in JPEG images.

It is intuitively clear, often demonstrated (e.g., in [15]), and theoretically studied[6] that longer secret messages ceteris paribus require more embedding changes and thus are statistically better detectable than smaller ones. Hence, capacity and embedding rate are related to security, the criterion to be discussed next.

2.3.2 Steganographic Security

The purpose of steganographic communication is to hide the mere existence of a secret message. Therefore, unlike cryptography, the security of a steganographic system is judged by the impossibility of detecting rather than by the difficulty of reading the message content. However, steganography builds on cryptographic principles for removing recognisable structure from message content, and to control information flows by the distribution of keys.

The steganalysis problem is essentially a decision problem (does a given object contain a secret message or not?), so decision-theoretic metrics qualify as measures of steganographic security and, by definition, equally as measures of steganalytic performance. In steganalysis, the decision maker is prone to two types of errors, for which the probabilities of occurrence are defined as follows (see also Table 2.1):

- The probability that the steganalyst fails to detect a stego object is called *missing probability* and is denoted by β.

[6] Capacity results can be found in [166] and [38] for specific memoryless channels, in Sect. 3 of [253] and [41] for stego systems defined on general artificial channels, and in [134] and [58] for stego systems with empirical covers. Theoretical studies of the trade-off between capacity and robustness are common (see, for example, [54, 172]), so it is surprising that the link between capacity and security (i.e., detectability) is less intensively studied.

- The probability that the steganalyst misclassifies a plain cover as a stego object is called *false positive probability* and denoted by α.

Further, $1 - \beta$ is referred to as *detection probability*. In the context of experimental observations of detector output, the term 'probability' is replaced by 'rate' to signal the relation to frequencies counted in a finite sample. In general, the higher the error probabilities, the better the security of a stego system (i.e., the worse the decisions a steganalyst makes).

Almost all systematic steganalysis methods do not directly come to a binary conclusion (cover or stego), but base their binary output on an internal state that is measured at a higher precision, for example, on a continuous scale. A decision threshold τ is used to quantise the internal state to a binary output. By adjusting τ, the error rates α and β can be traded off. A common way to visualise the characteristic relation between the two error rates when τ varies is the so-called *receiver operating characteristics* (ROC) curve. A typical ROC curve is depicted in Fig. 2.2 (a). It allows comparisons of the security of alternative stego systems for a fixed detector, or conversely, comparisons of detector performance for a fixed stego system. Theoretical ROC curves are always concave,[7] and a curve on the 45° line would signal perfect security. This means a detector performs no better than random guessing.

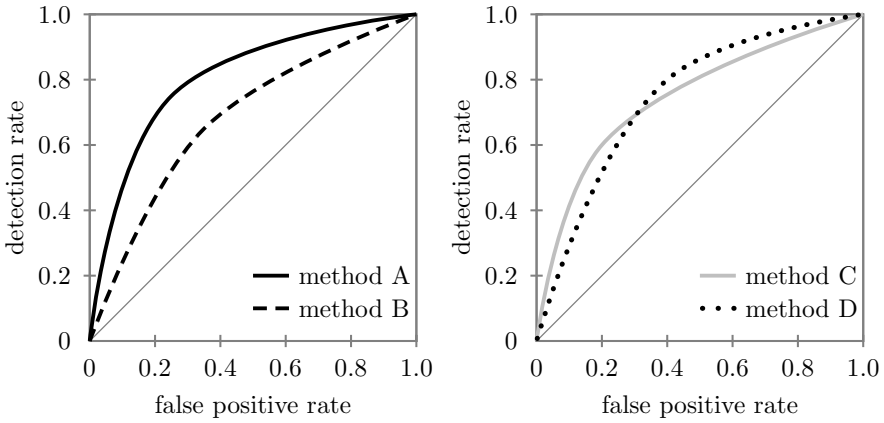
One problem of ROC curves is that they do not summarise steganographic security in a single figure. Even worse, the shape of ROC curves can be skewed so that the respective curves of two competing methods intersect (see Fig. 2.2 (b)). In this case it is particularly hard to compare different methods objectively.

As a remedy, many metrics derived from the ROC curve have been proposed to express steganographic security (or steganalysis performance) on a continuous scale, most prominently,

- the *detector reliability* as area under the curve (AUC), minus the triangle below the 45° line, scaled to the interval $[0, 1]$ (a measure of *insecurity*: values of 1 imply perfect detectability) [68],
- the false positive rate at 50% detection rate (denoted by FP_{50}),
- the *equal error rate* $\mathrm{EER} = \alpha \quad \Leftrightarrow \quad \alpha = \beta$,
- the *total minimal decision error* $\mathrm{TMDE} = \min_\tau \dfrac{\alpha + \beta}{2}$ [87], and
- the minimum of a cost- or utility-weighted sum of α and β whenever dependable weights are known for a particular application (for example, false positives are generally believed to be more costly in surveillance scenarios).

If one agrees to use one (and only one) of these metrics as the 'gold standard', then steganographic systems (or detectors) can be ranked according to its value, but statistical inference from finite samples remains tricky. A sort of inference test can be accomplished with critical values obtained from

[7] Estimated ROC curves from a finite sample of observations may deviate from this property unless a probabilistic quantiser is assumed to make the binary decision.



(a) univocal case (b) equivocal case

Fig. 2.2: ROC curve as measure of steganographic security. Left figure: stego system A is less secure than stego system B, because for any fixed false positive rate, the detection rate for A is higher than for B (in fact, both methods are insecure). Right figure: the relative (in)security of stego systems C and D depends on the steganalyst's decision threshold.

bootstrapping extensive simulation data, as demonstrated for a theoretical detector response in [235].

Among the list of ROC-based scalar metrics, there is no unique best option. Each metric suffers from specific weaknesses; for instance, AUC aggregates over practically irrelevant intervals of τ, EER and FP_{50} reflect the error rates for a single arbitrary τ, and the cost-based approach requires application-specific information.

As a remedy, recent research has tried to link theoretically founded metrics of statistical distinguishability, such as the Kullback–Leibler divergence between distributions of covers and stego objects, with practical detectors. This promises more consistent and sample-size-independent metrics of the amount of evidence (for the presence of a secret message) accumulated *per* stego object [127]. However, current proposals to approximate lower bounds (i.e., guaranteed insecurity) for typical stego detectors require thousands of measurements of the detector's internal state. So, more rapidly converging approximations from the machine learning community have been considered recently [188], but it is too early to tell if these metrics will become standard in the research community.

If the internal state is not available, a simple method to combine both error rates with an information-theoretic measure is the *binary relative entropy* of

two binary distributions with parameters $(\alpha, 1-\alpha)$ and $(1-\beta, \beta)$ [34]:

$$\mathsf{D}_{\mathsf{bin}}(\alpha,\beta) = \alpha \log_2 \frac{\alpha}{1-\beta} + (1-\alpha)\log_2 \frac{1-\alpha}{\beta}. \tag{2.2}$$

A value of $\mathsf{D}_{\mathsf{bin}}(\alpha,\beta) = 0$ indicates perfect security (against a specific decision rule, i.e., detector) and larger positive values imply better detectability. This metric has been proposed in the context of information-theoretic bounds for steganographic security. Thus, it is most useful to compare relatively secure systems (or weak detectors), but unfortunately it does not allow us to identify perfect separation ($\alpha = \beta = 0$). $\mathsf{D}_{\mathsf{bin}}(\alpha,\beta)$ converges to infinity as $\alpha, \beta \to 0$.

Finally and largely independently, *human perceptibility* of steganographic modifications in the cover media can also be subsumed to the security dimension, as demonstrated by the class of *visual attacks* [114, 238] against simple image steganography. However, compared to modern statistical methods, visual approaches are less reliable, depend on particular image characteristics, and cannot be fully automated. Note that in the area of watermarking, it is common to use the term *transparency* to describe visual imperceptibility of embedding changes. There, visual artefacts are not considered as a security threat, because the *existence* of hidden information is not a secret. The notion of security in watermarking is rather linked to the difficulty of *removing* a mark from the media object. This property is referred to as *robustness* in steganography and it has the same meaning in both steganographic and watermarking systems, but it is definitely more vital for the latter.

2.3.3 Robustness

The term robustness means the difficulty of removing hidden information from a stego object. While removal of secret data might not be a problem as serious as its detection, robustness is a desirable property when the communication channel is distorted by random errors (channel noise) or by systematic interference with the aim to prevent the use of steganography (see Sect. 2.5 below). Typical metrics for the robustness of steganographic algorithms are expressed in distortion classes, such as additive noise or geometric transformation. Within each class, the amount of distortion can be further specified with specific (e.g., parameters of the noise source) or generic (e.g., peak signal-to-noise ratio, PSNR) distortion measures. It must be noted that robustness has not received much attention so far in steganography research. We briefly mention it here for the sake of completeness. The few existing publications on this topic are either quite superficial, or extremely specific [236]. Nevertheless, robust steganography is a relevant building block for the construction of secure and effective censorship-resistant technologies [145].

2.3.4 Further Metrics

Some authors define additional metrics, such as *secrecy*, as the difficulty of extracting the message content [253]. We consider this beyond the scope of steganographic systems as the problem can be reduced to a confidentiality metric of the cryptographic system employed to encrypt a message prior to embedding (see [12] for a survey of such metrics). The computational *embedding complexity* and the *success rate*, i.e., the probability that a given message can be embedded in a particular cover at a given level of security and robustness, become relevant for advanced embedding functions that impose constraints on the permissible embedding distortion (see Sect. 2.8.2). Analogously, one can define the *detection complexity* as the computational effort required to achieve a given combination of error rates (α, β), although even a computationally unbounded steganalyst in general cannot reduce error rates arbitrarily for a finite number of observations. We are not aware of focused literature on detection complexity for practical steganalysis.

2.4 Paradigms for the Design of Steganographic Systems

The literature distinguishes between two alternative approaches to construct steganographic systems, which are henceforth referred to as *paradigms*.

2.4.1 Paradigm I: Modify with Caution

According to this paradigm, function Embed of a stego system takes as input cover data provided by the user who acts as sender, and embeds the message by modifying the cover. Following a general belief that fewer and smaller changes are less detectable (i.e., are more secure) than more and larger changes, those algorithms are designed to carefully preserve as many characteristics of the cover as possible.

Such distortion minimisation is a good heuristic in the absence of a more detailed cover model, but is not always optimal. To build a simple counterexample, consider as cover a stereo audio signal in a frequency domain representation. A hypothetical embedding function could attempt to shift the phase information of the frequency components, knowing that phase shifts are not audible to human perception and difficult to verify by a steganalyst who is unaware of the exact positioning of the microphones and sound sources in the recording environment. Embedding a secret message by shifting k phase coefficients in both channels randomly is obviously less secure than shifting $2k$ coefficients in both channels symmetrically, although the embedding distortion (measured in the number of cover symbols changed) is doubled. This is so

because humans can hear phase differences between two mixing sources, and a steganalyst could evaluate asymmetries between the two channels, which are atypical for natural audio signals.

Some practical algorithms have taken up this point and deliberately modify more parts of the cover in order to restore some statistical properties that are known to be analysed in steganalytic techniques (for example, OutGuess [198] or statistical restoration steganography [219, 220]). However, so far none of the actively preserving algorithms has successfully defeated targeted detectors that search for particular traces of active preservations (i.e., evaluate other statistics than the preserved ones). Some algorithms even turned out to be less secure than simpler embedding functions that do not use complicated preservation techniques (see [24, 76, 187, 215]). The crux is that it is difficult to change all symbols in a high-dimensional cover consistently, because the entirety of dependencies is unknown for empirical covers and cannot be inferred from a single realisation (cf. Sect. 3.1.3).

2.4.2 Paradigm II: Cover Generation

This paradigm is of a rather theoretical nature: its key idea is to replace the cover as input to the embedding function with one that is computer-generated by the embedding function. Since the cover is created entirely in the sender's trusted domain, the generation algorithm can be modified such that the secret message is already formed at the generation stage. This circumvents the problem of unknown interdependencies because the exact cover model is implicitly defined in the cover generating algorithm (see Fig. 2.3 and cf. artificial channels, Sect. 2.6.1).

The main shortcoming of this approach is the difficulty of conceiving plausible cover data that can be generated with (indeterministic) algorithms. Note that the fact that covers are computer-generated must be plausible in the communication context.[8] This might be true for a few mathematicians or artists who exchange colourful fractal images at high definition,[9] but is less so if supporters of the opposition in authoritarian states discover their passion for mathematics. Another possible idea to build a stego system following this paradigm is a renderer for photo-realistic still images or videos that contain indeterministic effects, such as fog or particle motion, which could be modulated by the secret message. The result would still be recognisable as computer-generated art (which may be plausible in some contexts), but its

[8] If the sender pretended that the covers are representations of reality, then one would face the same dilemma as in the first paradigm: the steganalyst could exploit imperfections of the generating algorithm in modelling the reality.

[9] Mandelsteg is a tool that seems to follow this paradigm, but it turns out that the fractal generation is not dependent on the secret message. ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/

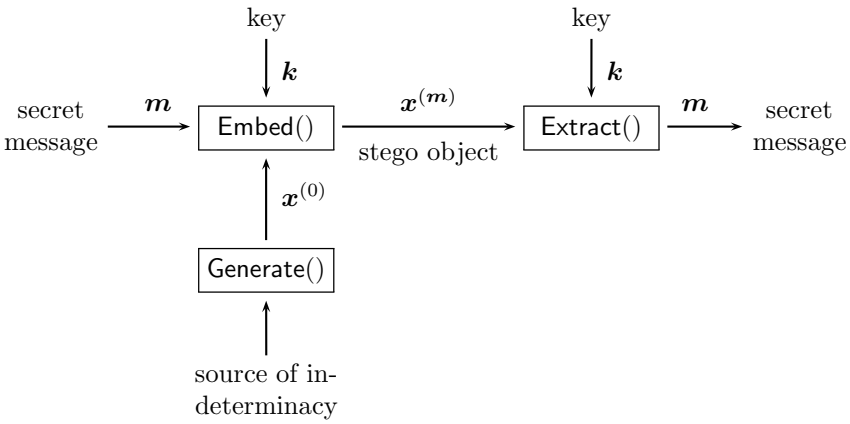


Fig. 2.3: Block diagram of stego system in the cover generation paradigm

statistical properties would not differ from similar art created with a random noise source to seed the indeterminism. Another case could be made for a steganographic digital synthesiser, which uses a noise source to generate drum and cymbal sounds.[10] Aside from the difficulty or high computational complexity of extracting such messages, it is obvious that the number of people dealing with such kind of media is much more limited than those sending digital photographs as e-mail attachments. So, the mere fact that uncommon data is exchanged may raise suspicion and thus thwart security. The only practical example of this paradigm we are aware of is a low-bandwidth channel in generated animation backgrounds for video conferencing applications, as recently proposed by Craver et al. [45].

A weaker form of this paradigm tries to avoid the plausibility problem without requiring consistent changes [64]. Instead of simulating a cover generation process, a plausible (ideally indeterministic, and at the least not invertible) cover *transformation* process is sought, such as downscaling or changing the colour depth of images, or, more general, lossy compression and redigitisation [65]. Figure 2.4 visualises the information flow in such a construction. We argue that stego systems simulating deterministic but not invertible transformation processes can be seen as those of paradigm I, 'Modify with Caution', with side information available exclusively to the sender. This is so because their security depends on the indeterminacy in the cover rather

[10] One caveat to bear in mind is that typical random number generators in creative software do not meet cryptographic standards and may in fact be predictable. Finding good pseudorandom numbers in computer-generated art may thus be an indication for the use of steganography. As a remedy, Craver et al. [45] call for 'cultural engineering' to make sending (strong) pseudorandom numbers more common.

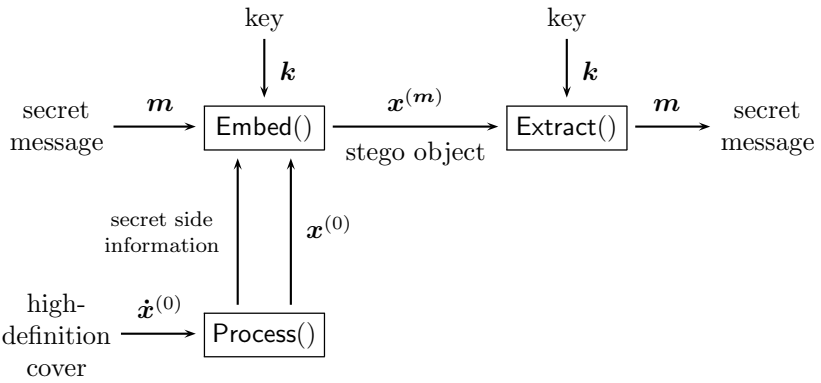


Fig. 2.4: Stego system with side information based on a lossy (or indeterministic) process: the sender obtains an information advantage over adversaries

than on artificially introduced indeterminacy (see Sect. 3.4.5 for further discussion of this distinction). Nevertheless, for the design of a stego system, the perspective of paradigm II may prove to be more practical: it is sometimes preferable for the steganographer to know precisely what the steganalyst most likely *will not know*, rather than to start with vague assumptions on what the steganalyst *might know*. Nevertheless, whenever the source of the cover is not fully under the sender's control, it is impossible to guarantee security properties because information leakage through channels unknown to the designer of the system cannot be ruled out.

2.4.3 Dominant Paradigm

The remainder of this chapter, in its function to provide the necessary background for the specific advances presented in the second part of this book, is confined to paradigm I, 'Modify with Caution'. This reflects the dominance of this paradigm in contemporary steganography and steganalysis research. Another reason for concentrating on the first paradigm is our focus on steganography and steganalysis in natural, that is empirical, covers. We argue in Sect. 2.6.1 that covers of (the narrow definition of) paradigm II constitute artificial channels, which are not empirical. Further, in the light of these arguments, we outline in Sect. 3.4.5 how the traditional distinction of paradigms in the literature can be replaced by a distinction of cover assumptions, namely (purely) empirical versus (partly) artificial cover sources.

2.5 Adversary Models

As in cryptography research, an adversary model is a set of assumptions defining the goals and limiting the computational power and knowledge of the steganalyst. Specifying adversary models is necessary because it is impossible to realise security goals against omnipotent adversaries. For example, if the steganalyst knows $\boldsymbol{x}^{(0)}$ for a specific act of communication, a secret message is detectable with probability $\mathsf{Prob}\left(i \neq 0 | \boldsymbol{x}^{(i)}\right) = 1 - 2^{-|\boldsymbol{m}|}$ by comparing objects $\boldsymbol{x}^{(i)}$ and $\boldsymbol{x}^{(0)}$ for identity. The components of an adversary model can be structured as follows:

- **Goals** The stego system is formulated as a probabilistic game between two or more competing players [117, for example].[11] The steganalyst's goal is to win this game, as determined by a utility function, with non-negligible probability. (A function $\mathsf{F} : \mathbb{Z}^+ \to [0,1]$ is called *negligible* if for every security parameter $\ell > 0$, for all sufficiently large y, $\mathsf{F}(y) < 1/y^\ell$).[12]
- **Computational power** The number of operations a steganalyst can perform and the available memory are bounded by a function of the security parameter ℓ, usually a polynomial in ℓ.
- **Knowledge** Knowledge of the steganalyst can be modelled as information sets, which may contain realisations of (random) variables as well as random functions ('oracles'), from which probability distributions can be derived through repeated queries (sampling).

From a security point of view, it is useful to define the strongest possible, but still realistic, adversary model. Without going into too many details, it is important to distinguish between two broad categories of adversary models: *passive* and *active* warden.[13]

2.5.1 Passive Warden

A passive warden is a steganalyst who does not interfere with the content on the communication channel, i.e., who has read-only access (see Fig. 2.5). The steganalyst's goal is to correctly identify the existence of secret messages by running function Detect (not part of the stego system, but possibly adapted to a specific one), which returns a metric to decide if a specific $\boldsymbol{x}^{(i)}$ is to be

[11] See Appendix E for an example game formulation (though some terminology is not introduced yet).

[12] Note that this definition does not limit the specification of goals to 'perfect' security (i.e., the stego system is broken if the detector is marginally better than random guessing). A simple construction that allows the specification of bounds to the error rates is a game in which the utility is cut down by the realisation of a random variable.

[13] We use the terms 'warden' and 'steganalyst' synonymously for steganographic adversaries. Other substitutes in the literature are 'attacker' and 'adversary'.

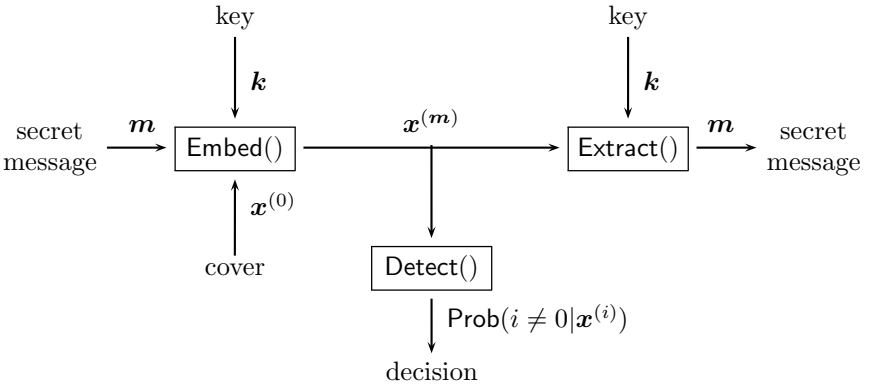


Fig. 2.5: Block diagram of steganographic system with passive warden

considered as a stego object or not. A rarely studied extension of this goal is to create evidence which allows the steganalyst to prove to a third party that steganography has been used.

Some special variants of the passive warden model are conceivable:

- Ker [123, 124] has introduced *pooled steganalysis*. In this scenario, the steganalyst inspects a set of suspect objects $\{\boldsymbol{x}_1^{(i_1)}, \dots, \boldsymbol{x}_N^{(i_N)}\}$ and has to decide whether steganography is used in any of them or not at all. This scenario corresponds to a situation where a storage device, on which secret data may be hidden in anticipation of a possible confiscation, is seized. In this setting, sender and recipient may be the same person. Research questions of interest deal with the strategies to distribute secret data in a batch of N covers, i.e., to find the least-detectable sequence $(i_1, \dots, i_N)$, as well as the optimal aggregation of evidence from N runs of Detect.
- Combining multiple outcomes of Detect is also relevant to *sequential steganalysis* of an infinite stream of objects $(\boldsymbol{x}_1^{(i_1)}, \boldsymbol{x}_2^{(i_2)}, \dots)$, pointed out by Ker [130]. Topics for study are, again, the optimal distribution $(i_1, i_2, \dots)$, ways to augment Detect by a memory of past observations $\mathsf{Detect}^{\mathcal{P}} : \mathfrak{P}(\mathcal{X}^*) \to \mathbb{R}$, and the timing decision about after how many observations sufficient evidence has accumulated.
- Franz and Pfitzmann [65] have studied, among other scenarios, the so-called *cover–stego-attacks*, in which the steganalyst has some knowledge $\hat{\boldsymbol{x}}^{(0)}$ about the cover of a specific act of communication, but not the exact realisation $\boldsymbol{x}^{(0)}$. This happens, for example, if a cover was scanned from a newspaper photograph: both sender and steganalyst possess an analogue copy, so the information advantage of the sender over the steganalyst is

merely the noise introduced in his private digitising process. Another example is embedding in MP3 files of commercially sold music.

- A more ambitious goal of a passive warden than detecting the presence of a secret message is learning its content. Fridrich et al. [84] discuss how the detector output for specific detectors can be used to identify likely stego keys.[14] This is relevant because the correct stego key cannot be found by exhaustive search if the message contains no recognisable redundancy, most likely due to prior encryption (with an independent crypto key). A two-step approach via the stego key can reduce the complexity of an exhaustive search for both stego and crypto keys from $\mathcal{O}(2^{2\ell})$ to $\mathcal{O}(2^{\ell+1})$ (assuming key sizes of ℓ bits each). Information-theoretic theorems on the secrecy of a message (as opposed to security $\leftrightarrow$ detectability) in a stego system can be found in [253].

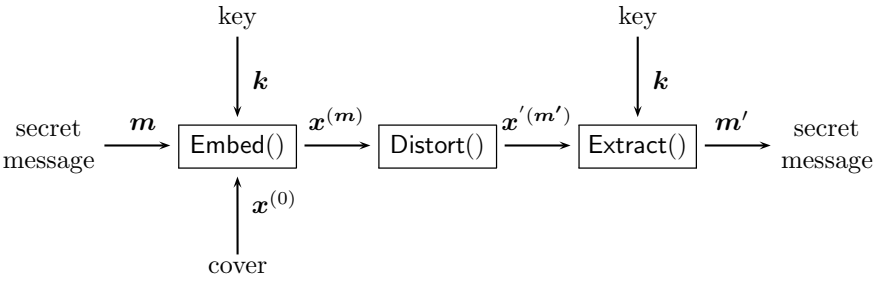


Fig. 2.6: Block diagram of steganographic system with active warden

2.5.2 Active Warden

In the active warden model, a steganalyst has read and write access to the communication channel. The wardens's goal is to prevent hidden communication or impede it by reducing the capacity of the hidden channel. This can be modelled by a distortion function $\mathsf{Distort} : \mathcal{X}^* \to \mathcal{X}^*$ in the communication channel (see Fig. 2.6). Note that systematic distortion with the aim to corrupt stego objects may also affect legitimate use of the communication channel adversely (e.g., by introducing visible noise or artefacts). Conversely, common transformations on legitimate channels may, as a side effect, distort

[14] We distinguish between 'stego' and 'crypto' keys only with regard to the secrecy of the message content: the former secures the fact that a message is present and the latter secures its content.

steganography despite not being designed with this intention (e.g., JPEG recompression or scaling on public photo communities or auction websites). Active warden models fit in the above-discussed structure for adversary models by specifying the warden's goals in a multistage game in which the options for the steganographers depend on previous moves of the warden.

Again, some variants of the active warden model are worth mentioning:

- A steganalyst, whose goal is to detect the use of steganography, could be in a position to supply the cover, or alter its value, before it is used as input to Embed by the sender. This happens, for example, when the steganalyst sells a modified digitisation device to the suspect sender, which embeds two watermarks in each output $\boldsymbol{x}^{(0)}$: one is robust against changes introduced by Embed and the other is fragile [155]. The use of steganography can be detected if an observed object $\boldsymbol{x}^{(i)}$ contains the robust watermark (which ensures that the tampered device has actually been used as the cover source), but not the fragile one (the indication that an embedding function as been applied on the cover). The robust watermark, which is a harder problem to realise, can be omitted if the fact that the cover is taken from the tampered device can be inferred from the context.
- A steganalyst can also actively participate as pretended communication partner in multiphase protocols, such as a covert exchange of a public stego key in public-key steganography (PKS). Consider a protocol where two communication partners perform a 'stego handshake' by first passing a public key embedded in a stego object $\boldsymbol{x}_1^{(\boldsymbol{k}_{\text{pub}})}$ from the sender (initiator) to the recipient, who uses it to encrypt a message that is returned in a stego object $\boldsymbol{x}_2^{(\mathsf{Encrypt}(\boldsymbol{m},\boldsymbol{k}_{\text{pub}}))}$. An active warden could act as initiator and 'challenge' a suspect recipient with a public-key stego object. The recipient can be convicted of using steganography if the reply contains an object from which a message with verifiable redundancy can be extracted using the respective private key. This is one reason why it is hard to build secure high capacity public-key steganography with reasonable cover assumptions[15] in the active warden model.

In practical applications we may face a combination of both passive and active adversaries. Ideal steganography thus should be a) secure to defeat passive steganalysis and b) robust to thwart attempts of interference with covert channels. This links the metrics discussed in Sect. 2.3 to the adversary models. The adversary model underlying the analyses in the second part of this book is the passive warden model.

[15] In particular, sampling cover symbols conditional on their history is inefficient. Such constructions have been studied by Ahn and Hopper [3] and an extension to adaptive active adversaries has been proposed by Backes and Cachin [8]. Both methods require a so-called 'rejection sampler'.

2.6 Embedding Domains

Before we drill down into the details of functions Embed and Extract in Sects. 2.7 and 2.8, respectively, let us recall the options for the domain of the cover representation $\mathcal{X}^*$. To simplify the notation, we consider covers $\mathcal{X}^n$ of finite dimension n.

2.6.1 Artificial Channels

Ahead of the discussion of empirical covers and their domains relevant to practical steganography, let us distinguish them from *artificial covers*. Artificial covers are sequences of elements x_i drawn from a *theoretically defined* probability distribution over a discrete channel alphabet of the underlying communication system. There is no uncertainty about the parameters of this distribution, nor about the validity of the cover model. The symbol generating process *is* the model. In fact, covers of the (strong form of) paradigm II, 'Cover Generation', are artificial covers (cf. Sect. 2.4).

We also use the term *artificial channel* to generalise from individual cover objects to the communication system's channel, which is assumed to transmit a sequence of artificial covers. However, a common simplification is to regard artificial covers of a single symbol, so the distinction between artificial channels and artificial covers can be blurry. Another simplification is quite common in theoretical work: a channel is called *memoryless* if there are no restrictions on what symbol occurs based on the history of channel symbols, i.e., all symbols in a sequence are independent. It is evident that memoryless channels are well tractable analytically, because no dependencies have to be taken into account.

Note that memoryless channels with known symbol distributions can be efficiently compressed to full entropy random bits and vice versa.[16] Random bits, in turn, are indistinguishable from arbitrary cipher text. In an environment where direct transmission of cipher text is possible and tolerated, there is no need for steganography. Therefore we deem artificial channels not relevant covers in practical steganography. Nevertheless, they do have a raison d'être in theoretical work, and we refer to them whenever we discuss results that are only valid for artificial channels.

The distinction between empirical covers and artificial channels resembles, but is not exactly the same as, the distinction between *structured* and *unstructured* covers made by Fisk et al. [60]. A similar distinction can also be found in [188], where our notion of artificial channels is called

[16] In theory, this also applies to stateful (as opposed to memoryless) artificial channels with the only difference being that the compression algorithm may become less efficient.

analytical model, as opposed to *high-dimensional model*, which corresponds to our notion empirical covers.[17]

2.6.2 Spatial and Time Domains

Empirical covers in spatial and time domain representations consist of elements x_i, which are discretised samples from measurements of analogue signals that are continuos functions of location (space) or time. For example, images in the spatial domain appear as a matrix of intensity (brightness) measurements sampled at an equidistant grid. Audio signals in the time domain are vectors of subsequent measurements of pressure, sampled at equidistant points in time (sampling rate). Digital video signals combine spatial and time dimensions and can be thought of as three-dimensional arrays of intensity measurements.

Typical embedding functions for the spatial or time domain modify individual sample values. Although small changes in the sample intensities or amplitudes barely cause perceptual differences for the cover as a whole, spatial domain steganography has to deal with the difficulty that spatially or temporally related samples are not independent. Moreover, these multivariate dependencies are usually non-stationary and thus hard to describe with statistical models. As a result, changing samples in the spatial or time domain consistently (i.e., preserving the dependence structure) is not trivial.

Another problem arises from file format conventions. From an information-theoretic point of view, interdependencies between samples are seen as a redundancy, which consumes excess storage and transmission resources. Therefore, common file formats employ lossy source coding to achieve leaner representations of media data. Steganography which is not robust to lossy coding would only be possible in uncompressed or losslessly compressed file formats. Since such formats are less common, their use by steganographers may raise suspicion and hence thwart the security of the covert communication [52].

2.6.3 Transformed Domain

A time-discrete signal $\boldsymbol{x} = (x_1, \ldots, x_n)$ can be thought of as a point in n-dimensional space $\mathbb{R}^n$ with a Euclidean base. The same signal can be expressed in an infinite number of alternative representations by changing the base. As long as the new base has at least rank n, this transformation is invertible and no information is lost. Different domains for cover representations are defined

[17] We do not follow this terminology because it confounds the number of dimensions with the empirical or theoretical nature of cover generating processes. We believe that although both aspects overlap often in practice, they should be separated conceptually.

by their linear transformation matrix $\boldsymbol{a}$: $\boldsymbol{x}_{\text{trans}} = \boldsymbol{a}\,\boldsymbol{x}_{\text{spatial}}$. For large n, it is possible to transform disjoint sub-vectors of fixed length from $\boldsymbol{x}$ separately, e.g., in blocks of $N^2 = 8 \times 8 = 64$ pixels for standard JPEG compression.

Typical embedding functions for the transformed domain modify individual elements of the transformed domain. These elements are often called 'coefficients' to distinguish them from 'samples' in the spatial domain.[18]

Orthogonal transformations, a special case, are rotations of the n-dimensional coordinate system. They are linear transformations defined by orthogonal square matrices, that is, $\boldsymbol{a}\,\boldsymbol{a}^{\mathsf{T}} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. A special property is that Euclidean distances in $\mathbb{R}^n$ space are invariant to orthogonal transformations. So, both embedding distortion and quantisation distortion resulting from lossy compression, measured as *mean square error* (MSE), are invariant to the domain in which the distortion is introduced.

Classes of orthogonal transformations can be distinguished by their ability to decorrelate elements of $\boldsymbol{x}$ if $\boldsymbol{x}$ is interpreted as a realisation of a random vector $\boldsymbol{X}$ with nonzero covariance between elements, or by their ability to concentrate the signal's energy in fewer (leading) elements of the transformed signal. The energy of a signal is defined as the square norm of the vector $e_{\boldsymbol{x}} = ||\boldsymbol{x}||$ (hence, energy is invariant to orthogonal transformations). However, both the optimal decorrelation transformation, the Mahalanobis transformation [208], as well as the optimal energy concentration transformation, the Karhunen–Loeve transformation [116, 158], also known as principal component analysis (PCA), are signal-dependent. This is impractical for embedding, as extra effort is required to ensure that the recipient can find out the exact transformation employed by the sender,[19] and not fast enough for the compression of individual signals. Therefore, good (but suboptimal) alternatives with fixed matrix $\boldsymbol{a}$ are used in practice.

The family of *discrete cosine transformations* (DCTs) is such a compromise, and thus it has a prominent place in image processing. A 1D DCT of column vector $\boldsymbol{x} = (x_1, \ldots, x_N)$ is defined as $\boldsymbol{y} = \boldsymbol{a}_{1\text{D}}\,\boldsymbol{x}$, with elements of the orthogonal matrix $\boldsymbol{a}_{1\text{D}}$ given as

$$a_{ij} = \sqrt{\frac{2}{N}} \cdot \cos\left(\frac{(2j-1)(i-1)\pi}{2N}\right)\left(1 + \frac{\delta_{i,1}}{2}(\sqrt{2}-2)\right), \quad 1 \le i, j \le N. \tag{2.3}$$

Operator $\delta_{i,j}$ is the Kronecker delta:

$$\delta_{i,j} = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j. \end{cases} \tag{2.4}$$

[18] We use 'sample' as a more general term when the domain does not matter.

[19] Another problem is that no correlation does not imply independence, which can be shown in a simple example. Consider the random variables $X = \sin\omega$ and $Y = \cos\omega$ with $\omega \sim \mathcal{U}_0^{2\pi}$; then, $\text{cor}(X, Y) \propto \mathsf{E}(XY) = \int_0^{2\pi} \sin u \cos u\, du = 0$, but X and Y are dependent, for example, because $\mathsf{Prob}(x = 0 \pm \varepsilon) < \mathsf{Prob}(x = 0 | y = 1) = 1/2$, $\varepsilon^2 \ll 1$. So, finding an uncorrelated embedding domain does not enable us to embed consistently with *all* possible dependencies between samples.

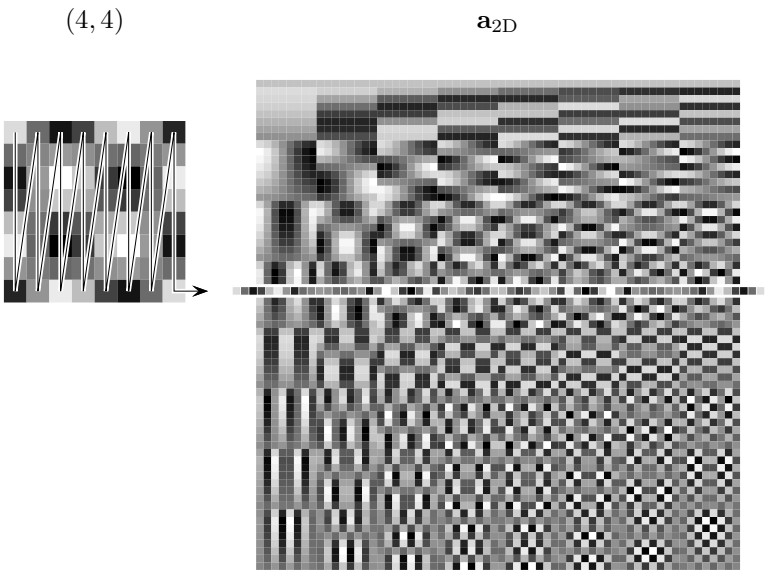


Fig. 2.7: 8×8 blockwise DCT: relation of 2D base vectors (example: subband $(4, 4)$) to row-wise representation in the transformation matrix $\boldsymbol{a}_{2D}$

Two 1D-DCT transformations can be combined to a linear-separable 2D-DCT transformation of square blocks with $N \times N$ elements. Let all k blocks of a signal $\boldsymbol{x}$ be serialised in columns of matrix $\boldsymbol{x}_{\boxplus}$; then,

$$\begin{aligned} \boldsymbol{y}_{\boxplus} &= \boldsymbol{a}_{2D}\, \boldsymbol{x}_{\boxplus} \quad \text{with} \\ \boldsymbol{a}_{2D} &= \big(\mathbf{1}_{N\times 1} \otimes \boldsymbol{a}_{1D} \otimes \mathbf{1}_{1\times N}\big) \odot \big(\mathbf{1}_{1\times N} \otimes \boldsymbol{a}_{1D} \otimes \mathbf{1}_{N\times 1}\big). \end{aligned} \tag{2.5}$$

Matrix $\boldsymbol{a}_{2D}$ is orthogonal and contains the N^2 base vectors of the transformed domain in rows. Figure 2.7 illustrates how the base vectors are represented in matrix $\boldsymbol{a}_{2D}$ and Fig. 2.8 shows the typical DCT base vectors visualised as 8×8 intensity maps to reflect the 2D character. The base vectors are arranged by increasing the horizontal and vertical spatial frequency subbands.[20] The upper-left base vector $(1, 1)$ is called the DC (direct current) component; all the others are AC (alternating current) subbands. Matrix $\boldsymbol{y}_{\boxplus}$ contains the transformed coefficients in rows, which serve as weights for the N^2 DCT base vectors to reconstruct the block in the inverse DCT (IDCT),

$$\boldsymbol{x}_{\boxplus} = \boldsymbol{a}_{2D}^{-1}\, \boldsymbol{y}_{\boxplus} = \boldsymbol{a}_{2D}^{\mathsf{T}}\, \boldsymbol{y}_{\boxplus}. \tag{2.6}$$

[20] Another common term for 'spatial frequency subband' is 'mode', e.g., in [189].

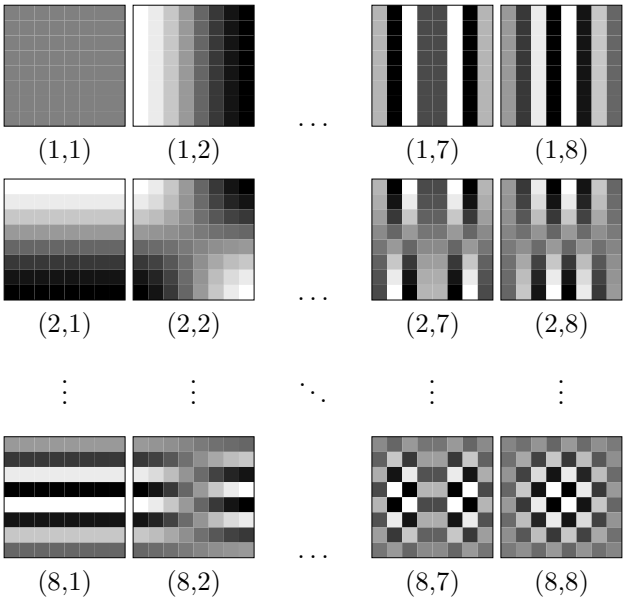


Fig. 2.8: Selected base vectors of 8×8 blockwise 2D DCT (vectors mapped to matrices)

In both $\boldsymbol{x}_{\boxplus}$ and $\boldsymbol{y}_{\boxplus}$, each column corresponds to one block. Note that a direct implementation of this mathematically elegant single transformation matrix method would require $\mathcal{O}(N^4)$ multiplication operations per block of $N \times N$ samples. Two subsequent 1D-DCT transformations require $\mathcal{O}(2N^3)$ operations, whereas *fast DCT* (FDCT) algorithms reduce the complexity further by factorisation and use of symmetries down to $\mathcal{O}(2N^2 - N \log_2 N - 2N)$ multiplications per block [57] (though this limit is only reachable at the cost of more additions, other trade-offs are possible as well).

Other common transformations not detailed here include the *discrete Fourier transformation* (DFT), which is less commonly used because the resulting coefficients contain phase information in the imaginary component of complex numbers, and the *discrete wavelet transformation* (DWT), which differs from the DCT in the base functions and the possibility to decompose a signal hierarchically at different scales.

In contrast to DCT and DFT domains, which are constructed from orthogonal base vectors, the *matching pursuit* (MP) 'domain' results from a decomposition with a highly redundant basis. Consequently, the decomposition is not unique and heuristic algorithms or other tricks, such as side information from related colour channels (e.g., in [35]), must be used to

ensure that both sender and recipient obtain the same decomposition path before and after embedding. Embedding functions operating in the MP domain, albeit barely tested with targeted detectors, are claimed to be more secure than spatial domain embedding because changes appear on a 'higher semantic level' [35, 36].

Unlike spatial domain representations in the special case of natural images, for which no general statistical model of the marginal distribution of intensity values is known, distributions of AC DCT coefficients tend to be unimodal and symmetric around 0, and their shape fits Laplace (or more generally, Student t and Generalised Gaussian) density functions reasonably well [148].

While orthogonal transformations between different domains are invertible in $\mathbb{R}^n$, the respective inverse transformation recovers the original values only approximately if the intermediate coefficients are rounded to fixed precision.[21] Embedding in the transformed domain, after possible rounding, is beneficial if this domain is also used on the channel, because subtle embedding changes are not at risk of being altered by later rounding in a different domain. Nevertheless, some stego systems intentionally choose a different embedding domain, and ensure robustness to later rounding errors with appropriate channel coding (e.g., embedding function YASS [218]).

In many lossy compression algorithms, different subbands are rescaled before rounding to reflect differences in perceptual sensitivity. Such scaling and subsequent rounding is called *quantisation*, and the scaling factors are referred to as *quantisation factors*. To ensure that embedding changes are not corrupted during quantisation, the embedding function is best applied on already quantised coefficients.

2.6.4 Selected Cover Formats: JPEG and MP3

In this section we review two specific cover formats, JPEG still images and MP3 audio, which are important for the specific results in Part II. Both formats are very popular (this is why they are suitable for steganography) and employ lossy compression to minimise file sizes while preserving good perceptual quality.

2.6.4.1 Essentials of JPEG Still Image Compression

The Joint Photographic Expert Group (JPEG) was established in 1986 with the objective to develop digital compression standards for continuous-tone still images, which resulted in ISO Standard 10918-1 [112, 183].

[21] This does not apply to the class of invertible integer approximations to popular transformations, such as (approximate) integer DCT and integer DWT; see, for example, [196].

Standard JPEG compression cuts a greyscale image into blocks of 8×8 pixels, which are separately transformed into the frequency domain by a 2D DCT. The resulting 64 DCT coefficients are divided by subband-specific quantisation factors, calculated from a JPEG quality parameter q, and then rounded to the closest integer. In the notation of Sect. 2.6.3, the quantised DCT coefficients $\boldsymbol{y}^*_{\boxplus}$ can be obtained as follows:

$$\boldsymbol{y}^*_{\boxplus} = \lfloor \overline{\boldsymbol{q}}\, \boldsymbol{y}_{\boxplus} + 1/2 \rfloor \quad \text{with} \quad \overline{q}_{i,j} = \begin{cases} (\mathsf{Quant}(q,i))^{-1} & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Function $\mathsf{Quant} : \mathbb{Z}^+ \times \{1, \ldots, 64\} \to \mathbb{Z}^+$ is publicly known and calculates subband-specific quantisation factors for a given JPEG compression quality q. The collection of 64 quantisation factors on the diagonal of $\boldsymbol{q}$ is often referred to as *quantisation matrix* (then aligned to dimensions 8×8). In general, higher frequency subbands are quantised with larger factors. Then, the already quantised coefficients are reordered in a zigzag manner (to cluster 0s in the high-frequency subbands) and further compressed by a lossless run-length and Huffman entropy [107] encoder. A block diagram of the JPEG compression process is depicted in Fig. 2.9.

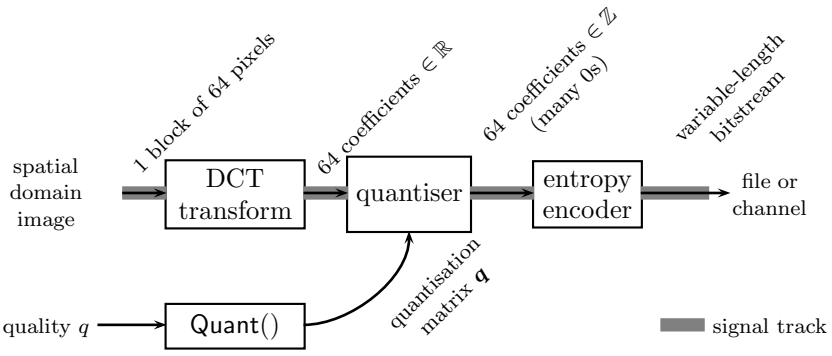


Fig. 2.9: Signal flow of JPEG compression (for a single colour component)

Colour images are first decomposed into a luminance component $\boldsymbol{y}$ (which is treated as a greyscale image) and two chrominance components $\boldsymbol{c}_R$ and $\boldsymbol{c}_B$ in the YCrCb colour model. The resolution of the chrominance components is usually reduced by factor 2 (owing to the reduced perceptibility of small colour differences of the human visual system) and then compressed separately in the same way as the luminance component. In general, the

chrominance components are quantised with larger factors than the luminance component.

All JPEG operations in Part II were conducted with `libjpeg`, the Independent JPEG Group's reference implementation [111], using default settings for the DCT method unless otherwise stated.

2.6.4.2 Essentials of MP3 Audio Compression

The Moving Picture Expert Group (MPEG) was formed in 1988 to produce standards for coded representations of digital audio and video. The popular MP3 file format for lossy compressed audio signals is specified in the ISO/MPEG 1 Audio Layer-3 standard [113]. A more scientific reference is the article by Brandenburg and Stoll [30].

The MP3 standard combines several techniques to maximise the trade-off between perceived audio quality and storage volume. Its main difference from many earlier and less efficient compression methods is its design as a two-track approach. The *first track* conveys the audio information, which is first passed to a filter bank and decomposed into 32 equally spaced frequency subbands. These components are separately transformed to the frequency domain with a *modulated discrete cosine transformation* (MDCT).[22] A subsequent quantisation operation reduces the precision of the MDCT coefficients. Note that the quantisation factors are called 'scale factors' in MP3 terminology. Unlike for JPEG compression, these factors are not constant over the entire stream. Finally, lossless entropy encoding of the quantised coefficients ensures a compact representation of MP3 audio data. The *second track* is a control track. Also, starting again from the *pulse code modulation* (PCM) input signal, a 1024-point FFT is used to feed the frequency spectrum of a short window in time as input to a psycho-acoustic model. This model emulates the particularities of human auditory perception, measures and values distortion, and derives masking functions for the input signal to cancel inaudible frequencies. The model controls the choice of block types and frequency band-specific scale factors in the first track. All in all, the two-track approach adaptively finds an optimal trade-off between data reduction and audible degradation for a given input signal. Figure 2.10 visualises the signal flow during MP3 compression.

Regarding the underlying data format, an MP3 stream consists of a series of *frames*. Synchronisation tags separate MP3 audio frames from other information sharing the same transmission or storage stream (e.g., video frames). For a given bit rate, all MP3 frames have a fixed compressed size and represent a fixed amount of 1,152 PCM samples. Usually, an MP3 frame contains 32 bits of header information, an optional 16 bit *cyclic redundancy check*

[22] The MDCT corresponds to the *modulated lapped transformation* (MLT), which transforms overlapping blocks to the frequency domain [165]. This reduces the formation of audible artefacts at block borders. The inverse transformation is accomplished in an overlap-add process.

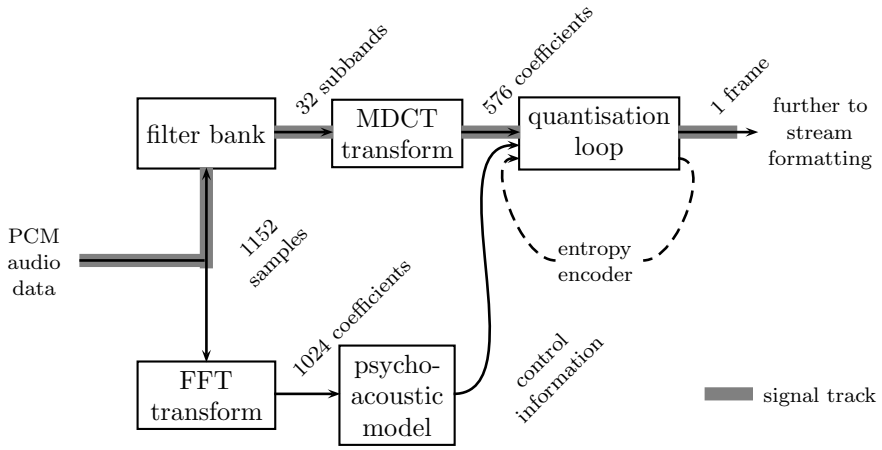


Fig. 2.10: Signal and control flow of MP3 compression (simplified)

(CRC) checksum, and two so-called *granules* of compressed audio data. Each granule contains one or two *blocks*, for mono and stereo signals, respectively. Both granules in a frame may share (part of) the scale factor information to economise on storage space. Since the actual block size depends on the amount of information that is required to describe the input signal, block and granule sizes may vary between frames. To balance the floating granule sizes across frames of fixed sizes efficiently, the MP3 standard introduces a so-called *reservoir* mechanism. Frames that do not use their full capacity are filled up (partly) with block data of subsequent frames. This method ensures that local highly dynamic sections in the input stream can be stored with over-average precision, while less demanding sections allocate under-average space. However, the extent of reservoir usage is limited in order to decrease the interdependencies between more distant frames and to facilitate resynchronisation at arbitrary positions in a stream. A schema of the granule-to-frame allocation in MP3 streams is depicted in Fig. 2.11.

2.6.5 Exotic Covers

Although the large majority of publications on steganography and steganalysis deal with digital representations of continuous signals as covers,

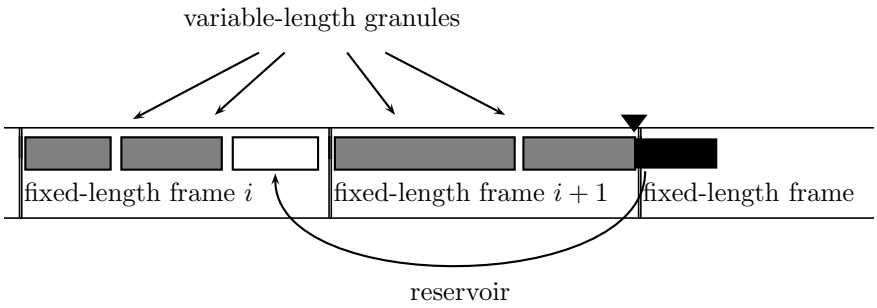


Fig. 2.11: MP3 stream format and reservoir mechanism

alternatives have been explored as well. We mention the most important ones only briefly.

Linguistic or *natural language* steganography hides secret messages in text corpuses. A recent literature survey [13] concludes that this branch of research is still in its infancy. This is somewhat surprising as text covers have been studied in the very early publications on mimic functions by Wayner [232], and various approaches (e.g., lexical, syntactic, ontologic or statistical methods) of automatic text processing are well researched in computer linguistics and machine translation [93].

Vector objects, meshes and general graph-structured data constitute another class of potential covers. Although we are not aware of specific proposals for steganographic applications, it is well conceivable to adapt principles from watermarking algorithms and increase (steganographic) security at the cost of reduced robustness for steganographic applications. Watermarking algorithms have been proposed for a large variety of host data, such as 2D vector data in digital maps [136], 3D meshes [11], CAD data [205], and even for very general data structures, such as XML documents and relational databases [92]. (We cite early references of each branch, not the latest refinements.)

2.7 Embedding Operations

In an attempt to give a modular presentation of design options for steganographic systems, we distinguish the high-level embedding function from low-level *embedding operations*.

Although in principle Embed may be an arbitrary function, in steganography it is almost universal practice to decompose the cover into *samples* and the secret message into bits (or q-ary symbols), and embed bits (or symbols) into samples independently. There are various reasons for this being so popular: ease of embedding and extracting, ability to use coding methods,

and ease of spreading the secret message over the cover. In the general setting, the assignment of message bits $m_j \in \{0,1\}$ to cover samples $x_i^{(0)}$ can be interleaved [43, 167]. Unless otherwise stated, we assume a pseudorandom permutation of samples using key $\boldsymbol{k}$ for secret-key steganography, although we abstract from this detail in our notation to improve readability. For embedding rates $p < 1$, random interleaving adds extra security by distributing the embedding positions over the entire cover, thus balancing embedding density and leaving the steganalyst uninformed about which samples have been changed for embedding (in a probabilistic sense). Below, in Sect. 2.8.2, we discuss alternative generalised interleaving methods that employ channel coding. These techniques allow us to minimise the number of changes, or to direct changes to specific parts of $\boldsymbol{x}^{(0)}$, the location of which remains a secret of the sender.

2.7.1 LSB Replacement

Least significant bit (LSB) replacement is probably the oldest embedding operation in digital steganography. It is based on the rationale that the rightmost (i.e., least significant) bit in digitised signals is so noisy that its bitplane can be replaced by a secret message imperceptibly:

$$x_i^{(1)} \leftarrow 2 \cdot \lfloor x_i^{(0)}/2 \rfloor + m_j. \tag{2.8}$$

For instance, Fig. 2.12 shows an example greyscale image and its (amplified) signal of the spatial domain LSB plane. The LSB plane looks purely random and is thus indistinguishable from the LSB plane of a stegotext with 12.5% secret message content. However, this impression is misleading as LSBs, despite being superficially noisy, are generally not independent of higher bitplanes. This empirical fact has led to a string of powerful detectors for LSB replacement in the spatial domain [46, 48, 50, 73, 74, 82, 118, 122, 126, 133, 151, 160, 238, 252, 257] and in the DCT domain [152, 153, 238, 243, 244, 248, 251]. Note that some implementations of LSB replacement in the transformed domain skip coefficients with values $x^{(0)} \in \{0, +1\}$ to prevent perceptible artefacts from altering many 0s to values $+1$ (0s occur most frequently due to the unimodal distribution with 0 mode). For the same reason, other implementations exclude $x^{(0)} = 0$ and modify the embedding function to

$$x_i^{(1)} \leftarrow 2 \cdot \left\lfloor (x_i^{(0)} - k)/2 \right\rfloor + k + m_j \quad \text{with} \quad k = \begin{cases} 0 \text{ for } x_i^{(0)} < 0 \\ 1 \text{ for } x_i^{(0)} > 0. \end{cases} \tag{2.9}$$

Probably the shortest implementation of spatial domain LSB replacement steganography is a single line of PERL proposed by Ker [118, p. 99]:

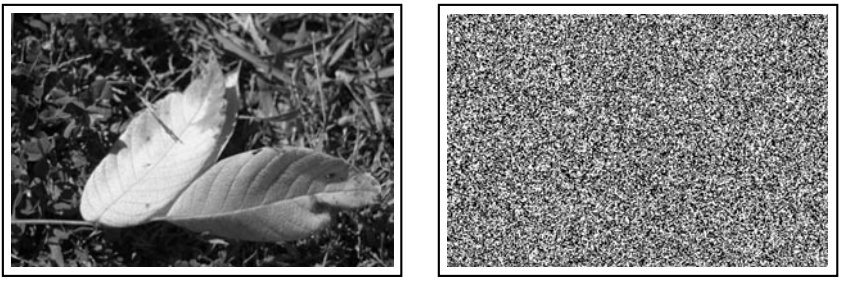


Fig. 2.12: Example eight-bit greyscale image taken from a digital camera and downsampled with nearest neighbour interpolation (left) and its least significant bitplane (right)

```
perl -n0777e '$_=unpack"b*",$_;split/(\s+)/,<STDIN>,5;
              @_[8]=~s{.}{$&&v254|chop()&v1}ge;
              print@_' <input.pgm >output.pgm secrettextfile
```

The simplicity of the embedding operation is often named as a reason for its practical relevance despite its comparative insecurity. Miscreants, such as corporate insiders, terrorists or criminals, may resort to manually typed LSB replacement because they must fear that their computers are monitored so that programs for more elaborate and secure embedding techniques are suspicious or risk detection as malware by intrusion detection systems (IDSs) [118].

2.7.2 LSB Matching (± 1)

LSB matching, first proposed by Sharp [214], is almost as simple to implement as LSB replacement, but much more difficult to detect in spatial domain images [121]. In contrast to LSB replacement, in which even values are never decremented and odd values never incremented,[23] LSB matching chooses the change for each sample x_i independently of its parity (and sign), for example, by randomising the sign of the change,

$$x_i^{(1)} \leftarrow x_i^{(0)} + \mathsf{LSB}(x_i^{(0)} - m_j) \cdot R_i \quad \text{with} \quad \frac{R_i + 1}{2} \sim \ddot{\mathcal{U}}_0^1. \tag{2.10}$$

Function $\mathsf{LSB} : \mathcal{X} \to \{0, 1\}$ returns the least significant bit of its argument,

[23] This statement ignores other conditions, such as in Eq. (2.9), which complicate the rule but do not solve the problem of LSB replacement that the steganalyst can infer the sign of potential embedding changes.

$$\mathsf{LSB}(x) = x - 2 \cdot \lfloor x/2 \rfloor = \mathsf{Mod}(x, 2). \tag{2.11}$$

R_i is a discrete random variable with two possible realisations $\{-1, +1\}$ that each occur with 50% probability. This is why LSB matching is also known as ± 1 embedding ('plus-minus-one', also abbreviated PM1). The random signs of the embedding changes avoid structural dependencies between the direction of change and the parity of the sample, which defeats those detection strategies that made LSB replacement very vulnerable. Nevertheless, LSB matching preserves all other desirable properties of LSB replacement. Message extraction, for example, works exactly in the same way as before: the recipient just interprets $\mathsf{LSB}(x_i^{(1)})$ as message bits.

If Eq. (2.10) is applied strictly, then elements $x_i^{(1)}$ may exceed the domain of $\mathcal{X}$ if $x_i^{(0)}$ is saturated.[24] To correct for this, $\boldsymbol{R}$ is adjusted as follows: $R_i = +1$ for $x_i^{(0)} = \inf \mathcal{X}$, and $R_i = -1$ for $x_i^{(0)} = \sup \mathcal{X}$. This does not affect the steganographic semantic for the recipient, but LSB matching reduces to LSB replacement for saturated pixels. This is why LSB matching is not as secure in covers with large areas of saturation. A very short PERL implementation for random LSB matching is given in [121].

Several variants of embedding functions based on LSB matching have been proposed in the literature and shall be recalled briefly:

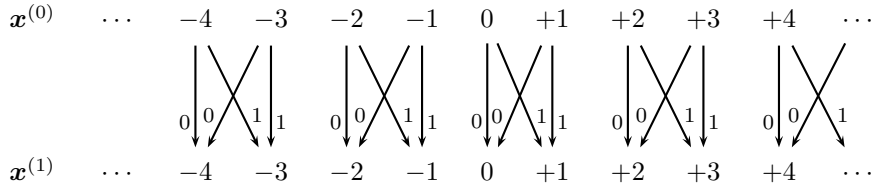
- **Embedding changes with moderated sign** If reasonably good distribution models are known for cover signals, then the sign of R_i can be chosen based on these models to avoid atypical deformation of the histogram. In particular, R_i should take value $+1$ with higher probability in regions where the density function has a positive first derivative, whereas $R_i = -1$ is preferable if the first derivative of the density function is negative. For example, the F5 algorithm [233] defines fixed signs of R_i depending on which side of the theoretical (0 mean) distribution of quantised JPEG AC coefficients a realisation $x_i^{(0)}$ is located. Hence, it embeds bits into coefficients by never increasing their absolute value.[25] Possible ambiguities in the steganographic semantic for the recipient can be dealt with by re-embedding (which gives rise to the 'shrinkage' phenomenon: for instance, algorithm F5 changes 50% of $x_i^{(0)} \in \{-1, +1\}$ without embedding a message bit [233]), or preferably by suitable encoding to avoid such cases preemptively (cf. Sect. 2.8.2 below).

[24] Saturation means that the original signal went beyond the bounds of $\mathcal{X}$. The resulting samples are set to extreme values $\inf \mathcal{X}$ or $\sup \mathcal{X}$.

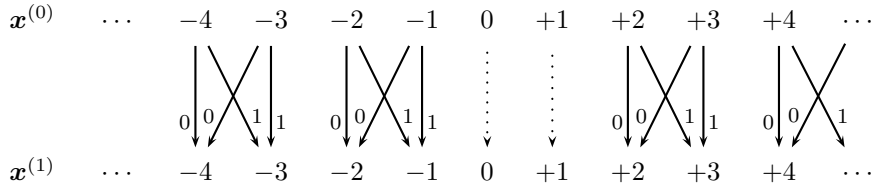
[25] Interestingly, while this embedding operation creates a bias towards 0 and thus changes the shape of the histogram, Fridrich and Kodowsky [86] have proven that this operation introduces the least overall embedding distortion if the unquantised coefficients are unknown (i.e., if the cover is already JPEG-compressed). This finding also highlights that small distortion and histogram preservation are competing objectives, which cannot be optimised at the same time.

- **Determining the sign of R_i from side information** Side information is additional information about the cover $\boldsymbol{x}^{(0)}$ available *exclusively* to the sender, whereas moderated sign embedding uses global rules or information shared with the communication partners. In this sense, side information gives the sender an advantage which can be exploited in the embedding function to improve undetectability. It is typically available when Embed goes along with information loss, for example, through scale reduction, bit-depth conversions [91], or JPEG (double-)compression [83] (cf. Fig. 2.4 in Sect. 2.4.2, where the lossy operation is explicit in function Process). In all these cases, $\boldsymbol{x}^{(0)}$ is available at high (real) precision and later rounded to lower (integer) precision. If R_i is set to the opposite sign of the rounding error, a technique known as *perturbed quantisation* (PQ), then the total distortion of rounding and embedding decreases relative to the independent case, because embedding changes always offset a fraction of the rounding error (otherwise, the square errors of both distortions are additive, a corollary of the theorem on sums of independent random variables). Less distortion is believed to result in less detectable stego objects, though this assumption is hard to prove in general, and pathologic counterexamples are easy to find.
- **Ternary symbols: determining the sign of R_i from the secret message** The direction of the change can also be used to convey additional information if samples of $\boldsymbol{x}^{(1)}$ are interpreted as ternary symbols (i.e., as representatives of $\mathbb{Z}_3$) [169]. In a fully ternary framework, a net capacity of $\log_2 3 \approx 1.585$ bits per cover symbol is achievable, though it comes at a cost of potentially higher detectabilily because now $2/3$ of the symbols have to be changed on average, instead of $1/2$ in the binary case (always assuming maximum embedding rates) [91]. A compromise that uses ternary symbols to embed one extra bit per block—the operation is combined with block codes—while maintaining the average fraction of changed symbols at $1/2$ has been proposed by Zhang et al. [254]. Ternary symbols also require some extra effort to deal with $x_i^{(0)}$ at the margins of domain $\mathcal{X}$.

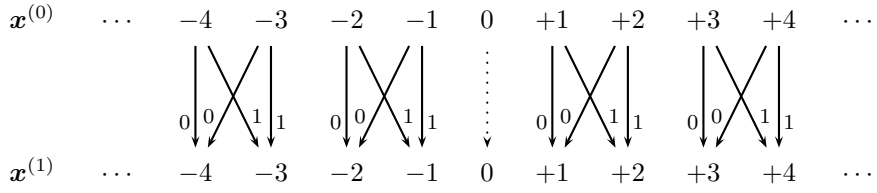
All embedding operations discussed so far have in common the property that the maximal absolute difference between individual cover symbols $x_i^{(0)}$ and their respective stego symbols $x_i^{(1)}$ is $1 \geq |x_i^{(0)} - x_i^{(1)}|$. In other words, the maximal absolute difference is minimal. A visual comparison of the similarities and differences of the mapping between cover and stego samples is provided in Fig. 2.13 (p. 44).



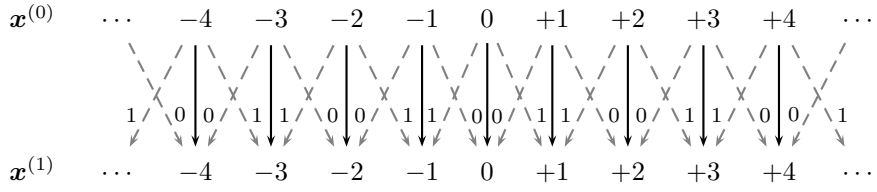
(a) Standard LSB replacement, Eq. (2.8)



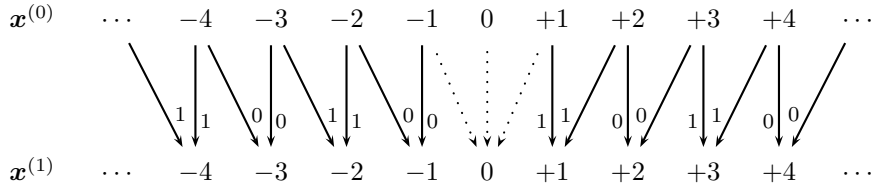
(b) LSB replacement, some values omitted (here: JSteg operation)



(c) LSB replacement, values omitted and shifted, Eq. (2.9)



(d) Standard LSB matching, Eq. (2.10)



(e) LSB matching, embedding changes with moderated sign (here: F5)

Fig. 2.13: Options for embedding operations with minimal maximum absolute embedding distortion per sample: $\max |x_i^{(0)} - x_i^{(1)}| = 1$; dotted arrows represent omitted samples, dashed arrows are options taken with conditional probability below 1 (condition on the message bit); arrow labels indicate steganographic semantic after embedding

2.7.3 Mod-k Replacement, Mod-k Matching, and Generalisations

If stronger embedding distortions $|x_i^{(0)} - x_i^{(1)}|$ than 1 are acceptable, then embedding operations based on both replacement and matching can be generalised to larger alphabets by dividing domain $\mathcal{X}$ into N disjoint sets of subsequent values $\{\mathcal{X}_i \mid \mathcal{X}_i \subset \mathcal{X} \wedge |\mathcal{X}_i| \geq k, 1 \leq i \leq N\}$. The steganographic semantic of each of the k symbols in the (appropriately chosen) message alphabet can be assigned to exactly one element of each subset $\mathcal{X}_i$. Such subsets are also referred to as *low-precision bins* [206].

For $\mathbb{Z}_{Nk} \subset \mathcal{X}$, a suitable breakdown is $\mathcal{X}_i = \{x \mid \lfloor x/k \rfloor = i - 1\}$ so that each $\mathcal{X}_i$ contains distinct representatives of $\mathbb{Z}_k$. The k symbols of the message alphabet are assigned to values of $x^{(1)}$ so that $\mathsf{Mod}(x^{(1)}, k) = m$. Mod-$k$ replacement maintains the low-precision bin after embedding (hence $x^{(0)}, x^{(1)} \in \mathcal{X}_i$) and sets

$$x_i^{(1)} \leftarrow k \cdot \lfloor x_i^{(0)}/k \rfloor + m_j. \tag{2.12}$$

For $k = 2^z$ with z integer, mod-k replacements corresponds to LSB replacement in the z least significant bitplanes.

Mod-k matching picks representatives of $m_j \equiv x_i^{(1)} (\text{mod } k)$ so that the embedding distortion $|x^{(0)} - x^{(1)}|$ is minimal (random assignment can be used if two suitable representatives are equally distant from the cover symbol $x^{(0)}$).

Further generalisations are possible if the low-precision bins have different cardinalities, for example, reflecting different tolerable embedding distortions in different regions of $\mathcal{X}$. Then, the message has to be encoded to a mixed alphabet. Another option is the adjustment of marginal symbol probabilities using *mimic functions*, a concept introduced by Wayner [232]. Sallee [206] proposed arithmetic decoders [240] as tools to build mimic functions that allow the adjustment of symbol probabilities in mod-k replacement conditionally on the low-precision bin of $x^{(0)}$.

Figure 2.14 illustrates the analogy between source coding techniques and mimic functions: in traditional source coding, function Encode compresses a nonuniformly distributed sequence of source symbols into a, on average, shorter sequence of uniform symbol distribution. The original sequence can be recovered by Decode with side information about the source distribution. Mimic functions useful in steganography can be created by swapping the order of calls to Encode and Decode: a uniform message sequence can be transcoded by Decode to an exogenous target distribution (most likely to match or 'mimic' some statistical property of the cover), whereas Encode is called at the recipient's side to obtain the (uniform, encrypted) secret message sequence.

Stochastic modulation embedding [72] is yet another generalisation of mod-k matching which allows (almost) arbitrary distribution functions for the

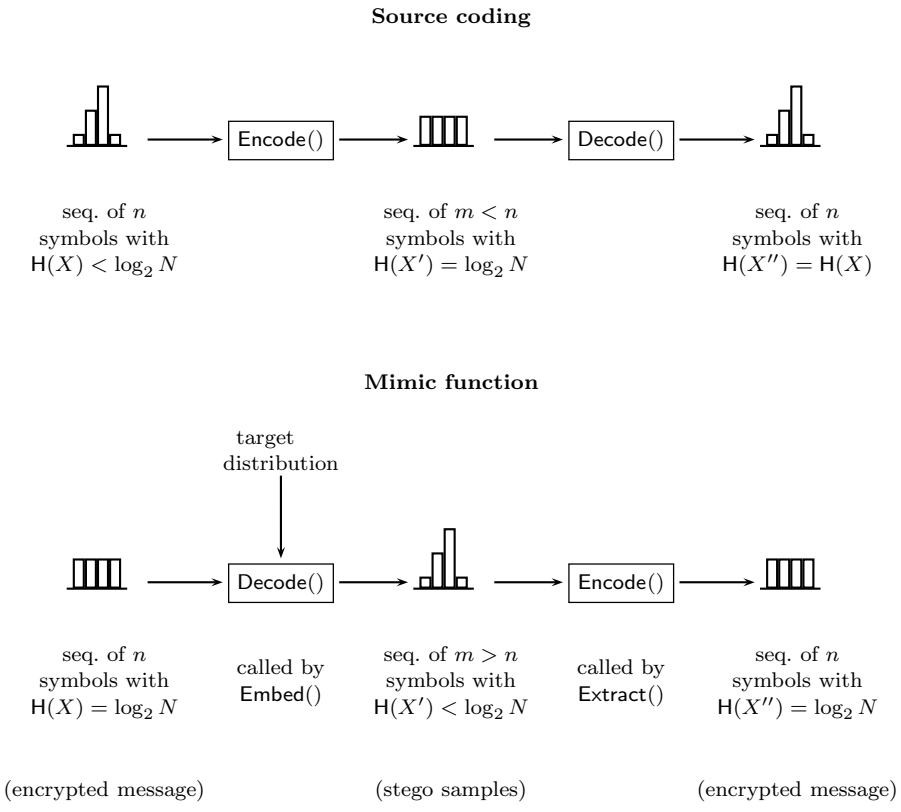


Fig. 2.14: Application of source coding techniques for entropy encoding (top) and as mimic function for embedding (bottom). The alphabet size is N and input sequences are identical to output sequences in both cases

random variable R in Eq. (2.10). The sender uses a pseudorandom number generator (PRNG) with a seed derived from the secret key to draw realisations from R_i. This ensures that the recipient can reproduce the actual sequence of r_i and determine the positions of samples where $|r_i|$ is large enough so that both steganographic message bits could be embedded by either adding or subtracting r_i from $x_i^{(0)}$ to obtain $x_i^{(1)}$. Extract evaluates only these 'usable' positions while skipping all others.

Finally, *spread spectrum image steganography* (SSIS) [167] can be seen as an approximate version of stochastic modulation (though invented before) which does not preemptively skip unusable realisations of R_i. To achieve comparable embedding capacities, on average higher embedding distortions

have to be accepted, which require extra redundancy through error correction codes and signal restoration techniques on the recipient's side. However, this extra effort lends SSIS a slight advantage over pure stochastic modulation in terms of robustness. SSIS, despite its name, is not limited to images as cover.

2.7.4 Multi-Sample Rules

As it is difficult to ensure that samples can be modified independently without leaving detectable traces, multi-sample rules have been proposed to change samples $x_i^{(0)}$ conditional on the realisations of other samples $x_j^{(0)}, j \neq i$, or even jointly. We distinguish broadly between two kinds of reference samples:

- Reference samples $x_j^{(0)}$ can be located in either spatial or temporal proximity, where the dependencies are assumed to be stronger than between more distant samples.
- Aggregate information of all samples in a cover object can serve as reference information. The idea here is to preserve macroscopic statistics of the cover.

One example for the first kind is the embedding operation of the CAS scheme by Lou and Sung [159], which evaluates the average intensity of the top-left adjacent pixels as well as the bottom-right adjacent pixels to calculate the intensity of the centre pixel conditional on the (encrypted) message bit (we omit the details for brevity). However, the CAS scheme shares a problem of multi-sample rules which, if not carefully designed, often ignore the possibility that a steganalyst who knows the embedding relations between samples can count the number of occurrences in which these relation hold exactly. This information, possibly combined with an analysis of the distribution of the exact matches, is enough to successfully detect the existence of hidden messages [21]. Another caveat of this kind of multi-sample rule is the need to ensure that subsequent embedding changes to the reference samples do not wreck the recipient's ability to identify the embedding positions (i.e., the criterion should be invariant to embedding operations on the reference samples).

Pixel-value differencing (PVD) in spatial domain images is another example of the first kind. Here, mod-k replacement is applied to intensity differences between pairs [241] or tuples [39] of neighbouring samples, possibly combined with other embedding operations on intensity levels or compensation rules to avoid unacceptable visible distortion [242]. Zhang and Wang [256] have proposed a targeted detector for PVD.

Examples for the second kind of multi-sample rules are OutGuess by Provos [198] and StegHide by Hetzl and Mutzel [102]. OutGuess employs LSB replacement in JPEG DCT coefficients, but flips additional correction LSBs to preserve the marginal histogram distributions. This increases the

average distortion per message bit and makes the stego system less secure against all kinds of detectors which do not only rely on marginal distributions. For instance, the detector by Fridrich et al. [76], calculates blockiness measures in the spatial domain. StegHide [102] preserves marginal distributions of arbitrary covers by exchanging positions of elements in $\boldsymbol{x}^{(0)}$ rather than altering values independently. A combinatorial solution is found by expressing the relations for possible exchanges as edges of a (possibly weighted) graph, which is solved by maximum cardinality matching. Successful steganalysis of StegHide has been reported for audio [204] and JPEG [157] covers. Both detectors evaluate statistics beyond the preserved marginal distributions.

2.7.5 Adaptive Embedding

Adaptive embedding can be seen as a special case of multi-sample rules; however, information from reference samples is not primarily used to apply consistent changes, but rather to identify locations where the distortion of single-sample embedding operations is least detectable. The aim is to concentrate the bulk of necessary changes there. Adaptive embedding can be combined with most of the above-discussed embedding operations. Ideally, the probability that the embedding operation does not modify a particular sample value should be proportional to the information advantage of the steganalyst from observing this particular sample in a modified realisation[26]:

$$\mathsf{Prob}(x_i^{(1)} = x_i^{(0)}) \propto \mathsf{Prob}(j \neq 0 | \boldsymbol{x}^{(j)} \wedge x_i^{(j)} \neq x_i^{(0)}) - \mathsf{Prob}(j \neq 0 | \boldsymbol{x}^{(j)}). \quad (2.13)$$

Unfortunately, the probabilities on the right-hand side of this relation are unknown in general (unless specific and unrealistic assumptions for the cover are made). Nevertheless, heuristic proposals for adaptive embedding rules are abundant for image steganography.[27] Lie and Chang [154] employ a model of the human visual system to control the number k of LSB planes used for mod-2^k replacement. Franz, Jerichow, Möller, Pfitzmann, and Stierand [63] exclude values close to saturation and close to the zero crossing of PCM digitised speech signals. Franz [62] excludes entire histogram bins from embedding based on the joint distribution with adjacent bins in a co-occurrence matrix built from spatial relations between pixels. Fridrich and Goljan [72]

[26] Note that this formulation states adaptive steganography as a local problem. Even if it could be solved for each sample individually, the solution would not necessarily be optimal on a global (i.e., cover-wide) scope. This is so because the individual information advantage may depend on other samples' realisations. In this sense, Eq. (2.13) is slightly imprecise.

[27] Despite the topical title 'Adaptive Steganography' and some (in our opinion) improper citations in the context of adaptive embedding operations, reference [37] does not deal with adaptive steganography according to this terminology. The paper uses adaptive in the sense of anticipating the steganalyst's exact detection method, which we deem rather unrealistic for security considerations.

discuss a content-dependent variant of their stochastic modulation operation, in which the standard deviation of the random variable R is modulated by an energy measure in the spatial neighbourhood. Similarly, adaptive ternary LSB matching is benchmarked against various other embedding operations in [91]. Aside from energy measures, typical image processing operators were suggested for adaptive steganography, such as dithering [66], texture [101] and edge detectors [180, 241, 242, 245].[28] Probably the simplest approach to adaptive steganography is due to Arjun et al. [6], who use the assumed perceptibility of intensity difference depending on the magnitude of $x_i^{(0)}$ as criterion, complemented by an exclusion of pixels with a constant intensity neighbourhood.

At first sight, adaptive embedding appears beneficial for the security of a stego system independent of the cover representation or embedding function [226] (at least if the underlying embedding operation is not insecure per se; so avoid LSB replacement). However, this only helps against myopic adversaries: one has to bear in mind that many of the adaptivity criteria are (approximately) invariant to embedding. In some embedding functions this is even a requirement to ensure correct extraction.[29] Adhering to Kerckhoffs' principle [135], this means that the steganalyst can re-recognise those regions where embedding changes are more concentrated. And in the worst case, the steganalyst could even compare statistics between the subset of samples which might have been affected from embedding and others that are most likely in their original state. Such kinds of detectors have been demonstrated against specific stego systems, for example, in [24]. More general implications of the game between steganographers and steganalysts on where to hide (and where to search, respectively) are largely unexplored. One reason for this gap might be the difficulty of quantifying the *detectability profile* [69] as a function of general cover properties. In Chapter 5 we present a method which is generally suitable to estimate cost functions for covers (and individual pixels, though not part of this book) empirically.

2.8 Protocols and Message Coding

This section deals with the architecture of stego systems on a more abstract level than the actual embedding operation on the signal processing layer. Topics of interest include the protocol layer, in particular assumptions on key distribution (Sect. 2.8.1), and options for coding the secret message to

[28] All these references evaluate the difference between neighbouring pixels to adjust k in mod-k replacement of the sample value or pairwise sample differences (i.e., PVDs). They differ in the exact calculation and correction rules to ensure that Extract works.

[29] Wet paper codes (cf. 2.8.2.2) have proved a recipe for correct extraction despite keeping the exact embedding positions a secret.

minimise the (detectability-weighted) distortion or leverage information advantages of the sender over the steganalyst (coding layer, Sect. 2.8.2).

2.8.1 Public-Key Steganography

In the context of steganography, the role of cryptography and of cryptographic keys in particular is to distinguish the communication partners from the rest of the world. Authorised recipients are allowed to recognise steganographic content and even extract it correctly, whereas third parties must not be able to tell stego objects apart from other communications. The common assumption in Simmons' initial formulation of the prisoners' problem [217] is that both communication partners share a common secret. This implies that both must have had the opportunity to communicate securely in the past to agree on a symmetric steganographic key. Moreover, they must have anticipated a situation in which steganographic communication is needed.[30]

Cryptography offers ways to circumvent this key distribution problem by using *asymmetric* cryptographic functions that operate with pairs of public and private keys. There exist no proposals like 'asymmetric steganography' for a direct analogy in steganography. Such a construction would require a trapdoor embedding function that is not invertible without the knowledge of a secret (or vast computational resources). However, by combining asymmetric cryptography with symmetric embedding functions, it is possible to construct so-called *public-key steganographic systems* (acronym PKS, as opposed to SKS for *secret-key steganography*).

The first proposal of steganography with public keys goes back to Anderson's talk on the first Information Hiding Workshop in 1996 [4]. Since, his work has been extended by more detailed considerations of active warden models [5]. The construction principles are visualised as a block diagram in Fig. 2.15, where we assume a passive warden adversary model. The secret message is encrypted with the public key of the recipient using an asymmetric cryptographic function, then (optionally) encoded so that encrypted message bits can be adapted to marginal distributions of the cover (mimic function) or placed in the least conspicuous positions in the cover. A keyless embedding function finally performs the actual embedding.[31] The recipient extracts a bitstream from each received object, feeds it to the decoder and subsequently tries to decrypt it with his or her private key. If the decryption

[30] It is obvious that allowing secret key exchanges in general when already 'locked in Simmons' prison' would weaken the assumptions on the communication restrictions: communication partners who are allowed to exchange keys (basically random numbers) can communicate *anything* through this channel.

[31] For example, a symmetric embedding function suitable for SKS with globally fixed key $\boldsymbol{k} = \text{const}$.

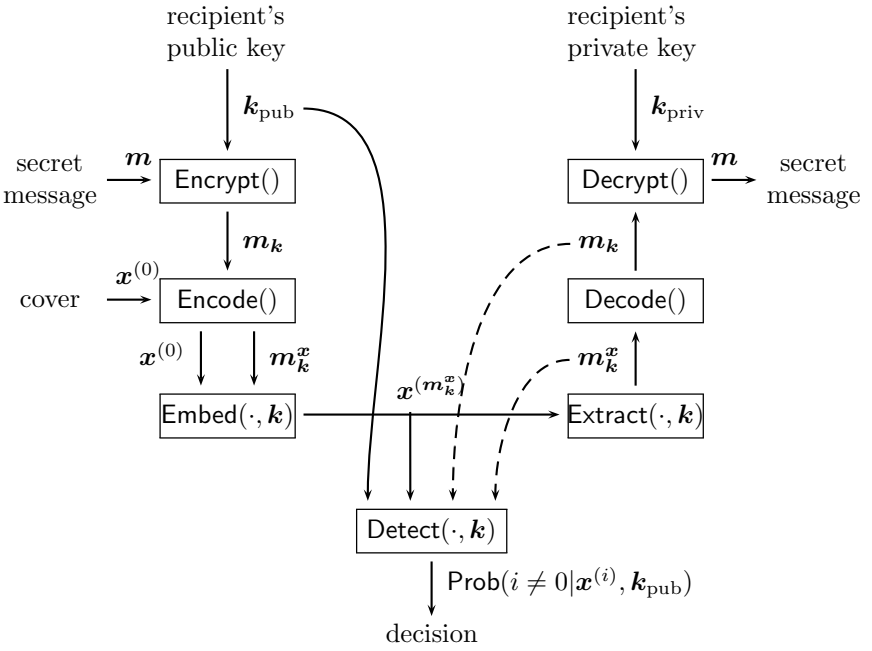


Fig. 2.15: Block diagram of public-key stego system with passive warden. Dashed lines denote that the information can be derived from $\boldsymbol{x}^{(\boldsymbol{m}_{\boldsymbol{k}}^{\boldsymbol{x}})}$ with public knowledge. The global 'key' $\boldsymbol{k}$ is optional and can be part of Embed, Extract and Detect (Kerckhoffs' principle)

succeeds, the recipient recognises that the received object was actually a stego object and retrieves the secret message.[32]

It is obvious that such a PKS system can never be more secure than the underlying SKS stego system consisting of Embed and Extract for random messages of length $|\boldsymbol{m}|$. In addition, as can be seen from the high number of arrows pointing to the steganalyst's function Detect, it is important for the security of the construction that none of

- the stego object $\boldsymbol{x}^{(\boldsymbol{m}_{\boldsymbol{k}}^{\boldsymbol{x}})}$,
- the bitstream generated by the message encoder $\boldsymbol{m}_{\boldsymbol{k}}^{\boldsymbol{x}}$, and
- the encrypted message $\boldsymbol{m}_{\boldsymbol{k}}$

be statistically distinguishable between clean covers and stego objects, even with knowledge of the recipient's public key $\boldsymbol{k}_{\text{pub}}$ (and, if it exists, knowledge

[32] Note that the message coding is implicit as part of Embed in the original publication. The distinction is made in the figure to emphasise which components of the system must output information indistinguishable between clean covers and stego objects [16].

of the global 'key' $\boldsymbol{k}$ used in the symmetric stego system Embed and Extract). In other words, Extract applied to arbitrary covers must always return a random sequence (possibly correlated to $\boldsymbol{x}$, but never revealing that information about $\boldsymbol{x}^{(0)}$ if $\boldsymbol{x}^{(p)}$ with $p > 0$ has been transmitted). Moreover, Decode applied to any possible output of Extract should be indistinguishable from ciphertexts created with Encrypt and the recipient's public key $\boldsymbol{k}_{\text{pub}}$. Only few asymmetric encryption schemes produce pseudorandom ciphertexts (e.g., [171] for a scheme based on elliptic curves, which has the nice property that it produces shorter ciphertexts than RSA or Diffie–Hellman-based alternatives), and well-known number-theoretic schemes in $\mathbb{Z}_{\mathfrak{p}}$ or $\mathbb{Z}_{\mathfrak{n}}$, with $\mathfrak{p}$ prime and $\mathfrak{n}$ semi-prime, can be used for PKS only in conjunction with a *probabilistic bias removal* (PBR) procedure [246].[33]

Initiating a steganographic communication relation with public keys requires a key exchange protocol, in which the recipient transmits his or her public key to the sender (and thus, at the same time, reveals his or her intention to communicate covertly). Assuming that sending keys openly is considered as suspicious, the public key itself has to be embedded as a secret message [44]. Again, one has to ensure that public keys are pseudorandom, which is not the case for the RSA-based key exchange proposed by Craver [44] (because random numbers tend to have small factors, but the semi-prime $\mathfrak{n}$ part of the RSA public key does not).[34] Therefore, a Diffie–Hellman integer encryption scheme (DHIES) [2] augmented by a PBR for the key exchange should be sufficiently secure in the passive warden model (NB, against polynomial bounded adversaries; if secure SKS exists; if the hash and MAC functions in the concrete DHIES implementation are secure).

Steganographic key exchanges are yet more difficult in the active warden adversary model. As discussed before in Sect. 2.5.2 (p. 29), we are not aware of a solution to the 'stego challenge' problem. A different approach to completely avoid the troublesome key exchanges in PKS is the (convenient) assumption that all communication partners have access to a digital signature system and can reuse its keys for steganography [144].

Orthogonal to extensions of Anderson's construction [4, 21, 44, 94, 144], there are publications on public-key steganography originating from the cryptology community. This literature focuses on public-key steganographic systems with provable security properties even in active warden models [3, 8, 104, 150]. However, the cost of this formal rigour is practical irrelevance, essentially due to two constraints, namely unrealistic assumptions,

[33] This is so because valid ciphertexts $s < \mathfrak{n}$, but $\lceil \log_2 \mathfrak{n} \rceil$ bits are needed to store s, so the distribution of 0s and 1s in the most significant bit(s) is not uniform.

[34] One can differentiate between whether it is sufficient that a notably high number of clean covers 'contain' a plausible public key, or whether finding a cover that does not 'contain' a message distinguishable from possible public keys should be difficult. While the former condition seems reasonable in practice, the latter is stronger and allows an admittedly unrealistic regime in which all complying communication partners who 'have nothing to hide' actively avoid sending covers with plausible public stego keys in order to signal their 'stegophobia', and thus potential steganographers are singled out.

most importantly that cover symbols can be sampled from an artificial channel with a known distribution, and inefficiency (such as capacities of one bit per cover object). The difference between these rather theoretical constructions of provable secure steganography and practical systems are not specific to PKS and explained further in Sect. 3.4.4.

2.8.2 Maximising Embedding Efficiency

Another string of research pioneered by Anderson [4] and, more specifically, Crandall [43] and Bierbrauer [14] includes channel coding techniques in the embedding function to optimise the choice of embedding positions for minimal detectability. As soon as the length of the secret message to be embedded $|\boldsymbol{m}|$ is smaller than the number of symbols n in $\boldsymbol{x}^{(0)}$ (with binary steganographic semantic), the sender gains degrees of freedom on which symbols to change to embed $\boldsymbol{m}$ in the least-detectable way, that is, with highest *embedding efficiency*. In general, embedding efficiency η can be defined as the length of the secret message divided by a suitable distortion measure for the steganographic system and adversary model under consideration:

$$\eta = \frac{|\boldsymbol{m}|}{\text{embedding distortion}}. \tag{2.14}$$

We distinguish between two important specific distortion measures, although other metrics and combinations are conceivable as well.

2.8.2.1 Embedding Efficiency with Respect to the Number of Changes

A simple measure of distortion is the number of changes to cover $\boldsymbol{x}^{(0)}$ during embedding; hence, Eq. (2.14) can be written as

$$\eta_{\#} = \frac{|\boldsymbol{m}|}{\mathsf{D}_{\mathsf{H}}(\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(\boldsymbol{m})})} \qquad \text{with} \quad \mathsf{D}_{\mathsf{H}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_i (1 - \delta_{x_i y_i}) \,. \tag{2.15}$$

Function $\mathsf{D}_{\mathsf{H}} : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{Z}$ denotes the Hamming distance between two vectors of equal length. *Syndrome coding* is a technique borrowed from channel coding to improve $\eta_{\#}$ above a value of 2.[35] To cast our cover vectors (following optional key-dependent permutation) to the universe of block codes, we

[35] If bits in $\boldsymbol{m}$ and the steganographic semantic of symbols in $\boldsymbol{x}^{(0)}$ are independently distributed with maximum entropy, then on average one symbol has to be changed to embed two message bits (the steganographic semantic of cover symbols already matches the desired message bit with 50% chance).

interpret $\boldsymbol{x}^{(0)} = (x_1, \ldots, x_n) = \boldsymbol{x}_1^{(0)} || \boldsymbol{x}_2^{(0)} || \ldots || \boldsymbol{x}_{\lceil n/n_\square \rceil}^{(0)}$ as a concatenation of blocks of size $n_\square$ each. Let $\boldsymbol{d} \in \{0,1\}^*$ be an $l \times n_\square$ parity check matrix of a linear block code (with $\mathsf{rank}(\boldsymbol{d}) = l \leq n_\square$), and let $\boldsymbol{b}_j^{(0)} \in \{0,1\}^{n_\square}$ be the binary column vector of the steganographic semantic extracted from individual symbols of $\boldsymbol{x}_j^{(0)}$, the jth block of $\boldsymbol{x}^{(0)}$.

If the recipient, after extracting the steganographic semantic $\boldsymbol{b}_j^{(1)}$ from $\boldsymbol{x}_j^{(1)}$, always builds the matrix product

$$\boldsymbol{m}_j = \boldsymbol{d}\,\boldsymbol{b}_j^{(1)} \tag{2.16}$$

to decode l message bits $\boldsymbol{m}_j$, then the sender can rearrange Eq. (2.16) and search for the auxiliary vector $\boldsymbol{v}_j$ that solves Eq. (2.19) with minimal Hamming weight. Nonzero elements in $\boldsymbol{v}_j$ indicate $\mathsf{D_H}(\boldsymbol{v}, \boldsymbol{0})$ positions in $\boldsymbol{x}_j^{(0)}$ where the steganographic semantic has to be changed by applying the embedding operation,

$$\boldsymbol{v}_j = \boldsymbol{b}_j^{(1)} - \boldsymbol{b}_j^{(0)} \tag{2.17}$$

$$\boldsymbol{d}\,\boldsymbol{v}_j = \boldsymbol{d}\,\boldsymbol{b}_j^{(1)} - \boldsymbol{d}\,\boldsymbol{b}_j^{(0)} \tag{2.18}$$

$$\boldsymbol{d}\,\boldsymbol{v}_j = \boldsymbol{m}_j - \boldsymbol{d}\,\boldsymbol{b}_j^{(0)}. \tag{2.19}$$

The syndrome $\boldsymbol{d}\,\boldsymbol{b}_j^{(0)}$ lends its name to the technique.

Early proposals [43] for the creation of $\boldsymbol{d}$ suggest binary Hamming and Golay codes, which are both good error-correcting codes and covering codes (the latter is important for embedding purposes). All codes of the Hamming family [96] are perfect codes and share a minimum distance 3 and a covering radius 1, which implies that the weight of $\boldsymbol{v}_j$ never exceeds 1. The only remaining perfect binary code is the binary Golay code, which has minimum distance 7 and covering radius 3 [14]. The advantage of Hamming codes is that the search for $\boldsymbol{v}_j$ is computationally easy—it follows immediately from the difference between syndrome $\boldsymbol{d}\,\boldsymbol{b}_j^{(0)}$ and message $\boldsymbol{m}_j$. This is why Hamming codes, renamed as 'matrix codes' in the steganography community, found their way into practical embedding functions quickly [233, for example]. More recently, covering properties of other structured error-correcting codes, such as BCH [173, 210, 211, 250], Reed–Solomon [61], or simplex (for $|\boldsymbol{m}|/n$ close to 1) [88], as well as (unstructured) random linear codes [85], have been studied.

A common problem of structured error-correcting codes beyond the limited set of perfect codes are their comparatively weak covering properties and the exponential complexity (in $n_\square$) of the search for $\boldsymbol{v}_j$ with minimum weight (also referred to as *coset leader* in the literature). This imposes an upper limit on possible block size $n_\square$ and keeps the attainable embedding efficiencies $\eta_\#$ in the low region of the theoretical bound [14]. Even so, heuristics have been proposed to trade off computational and memory complexity, to employ

probabilistic processing, and to restrict the result set to approximate (local) solutions [71, 212]. More recent methods exploit structural properties of the code [250] or are based on *low-density generator matrix* (LDGM) codes. For the latter, approximate solutions can be found efficiently for very large $n_\square \approx n$ [71, 95]. LDGM solvers can handle weighted Hamming distances and seem to work with more general distortion measures (of which Sect. 2.8.2.2 is a special case).

Most coding techniques mentioned here are not limited to binary cases, and some generalisations to arbitrary finite fields exist (e.g., Bierbrauer [14] for the general theory, Willems and van Dijk [239] for ternary Hamming and Golay codes, Fridrich [69] for q-ary random codes on groups of binary samples, and Zhang et al. [255] for code concatenation of binary codes in 'layers').

2.8.2.2 Embedding Efficiency with Respect to the Severity of Changes

Consider a function that implements adaptive embedding (cf. Sect. 2.7.5), possibly taking into account additional side information,

$$\mathsf{Wet} : \mathcal{X}^n \times \{\mathbb{R}^n, \perp\} \to \{0,1\}^n, \tag{2.20}$$

which assigns each sample in $\boldsymbol{x}^{(0)}$ to one of two classes based on the severity of a change with respect to perceptibility or detectability. Samples that are safe to be changed are called 'dry' (value 0) and those that should not be altered are called 'wet' (value 1). A useful metaphor is a piece of paper besprinkled in rain, so that ink lasts only on its dry parts. After a while, primarily 'wet' and 'dry' regions cannot be told apart anymore. This led to the term *wet paper codes* for embedding, introduced by Fridrich, Goljan, and Soukal [83].

Possible denominators of Eq. (2.14) can be arbitrary projections of the value of Wet to a scalar, such as the number of 'wet' samples changed; or, if the co-domain of Wet is continuous, a weighted sum. For the sake of simplicity, we restrict the presentation to this (degenerated, but fairly common) binary case:

$$\eta_\circledcirc = \begin{cases} 1 \text{ for } \quad x_i^{(0)} = x_i^{(\boldsymbol{m})} \ \forall i \in \left\{ i \mid \mathsf{Wet}(\boldsymbol{x}^{(0)}, \cdot) = 1 \right\} \\ 0 \text{ otherwise.} \end{cases} \tag{2.21}$$

According to this definition, embedding is efficient if the message can be placed into the cover object without altering any 'wet' sample and the recipient is able to extract it correctly without knowing the value of Wet. A first proposal for this problem by Anderson [4] is known as *selection channel*: all elements of $\boldsymbol{x}^{(0)}$ are divided into $|\boldsymbol{m}| \ll n$ blocks $\boldsymbol{x}_1^{(0)} || \boldsymbol{x}_2^{(0)} || \ldots || \boldsymbol{x}_{|\boldsymbol{m}|}^{(0)}$. Then, the parity of the steganographic semantics of all samples in one block

is interpreted as a message bit. Only blocks for which the parity does not match the message bit, i.e., $m_i \neq \mathsf{Parity}(\boldsymbol{b}_i^{(0)})$, must be adjusted by selecting the least-detectable sample of $\boldsymbol{x}_i^{(0)}$ for the embedding operation. If $n/|\boldsymbol{m}|$ is sufficiently large and elements of $\boldsymbol{x}^{(0)}$ are assigned to blocks $\boldsymbol{x}_i^{(0)}$ randomly, then the probability that no 'wet' sample has to be changed is reasonably high.

The probability of successful embedding can be further improved by using *wet paper codes* (WPCs), a generalisation of the selection channel. As for the minimisation of the number of changes, block sizes $n_\square = |\boldsymbol{x}_i^{(0)}|$ are chosen larger (hundreds of samples) to accommodate l message bits per block. For each block, an $l \times n_\square$ parity check matrix $\boldsymbol{d}_j$ is populated using a pseudorandom number generator seeded with key $\boldsymbol{k}$. As before, $\boldsymbol{b}_j^{(0)}$ is the steganographic semantic extracted from $\boldsymbol{x}_j^{(0)}$, and $\overline{\boldsymbol{b}}_j^{(0)}$ is a decimated vector excluding all bits that correspond to 'wet' samples. Analogously, the respective columns in $\boldsymbol{d}_j$ are removed in the reduced $l \times (n_\square - k_j)$ matrix $\overline{\boldsymbol{d}}_j$ (k_j is the number of 'wet' samples in the jth block, and $n_\square - k_j \gtrapprox l$). Vector $\overline{\boldsymbol{v}}_j$ indicates the embedding positions after inserting 0s for the omitted 'wet' samples and can be obtained by solving this equation with the Gaussian elimination method over the finite field $\mathbb{Z}_2$:[36]

$$\overline{\boldsymbol{d}}_j\, \overline{\boldsymbol{v}}_j = \boldsymbol{m}_j - \boldsymbol{d}_j\, \boldsymbol{b}_j^{(0)}. \tag{2.22}$$

As shown in [31] (cited from [83]), solutions for this system exist with high probability if $\boldsymbol{d}_j$ is sparsely populated. Unlike in the case of minimal changes, *any* solution is sufficient and there are no constraints with regard to the Hamming weight of $\overline{\boldsymbol{v}}_j$. The decoding operation is similar to Eq. (2.16) and uses the unreduced random matrix $\boldsymbol{d}_j$, since the recipient by definition does not know which columns were dropped due to 'wet' samples:

$$\boldsymbol{m}_j = \boldsymbol{d}_j\, \boldsymbol{b}_j^{(1)}. \tag{2.23}$$

Detailed strategies to embed the dimension of $\boldsymbol{d}$ (needed by the recipient) as metadata (obviously not using WPC) as well as a computationally less complex substitute for the Gaussian elimination, which exploits a specific stochastic structure of row and column weights in $\boldsymbol{d}_j$ and $\overline{\boldsymbol{d}}_j$, can be found in [80] and [81].

[36] Wet paper codes can be generalised to finite fields $\mathbb{Z}_{2^k}$ if k bits are grouped to one symbol, or to arbitrary finite fields if the underlying cover domain $\mathcal{X}$ and embedding operations support q-ary symbols.

2.8.2.3 Summary

The gist of the sections on maximising embedding efficiency for the remainder of this book is twofold:

1. The actual gross message length may exceed twice the number of embedding changes.
2. For secret-key steganography[37] with sufficiently large n and ratio of secure embedding positions, appropriate codes exist to concentrate the embedding changes in arbitrary locations of $\boldsymbol{x}^{(0)}$ without the need to share knowledge about the embedding positions with the recipient.

Further details on coding in steganography are beyond the scope of this work.

2.9 Specific Detection Techniques

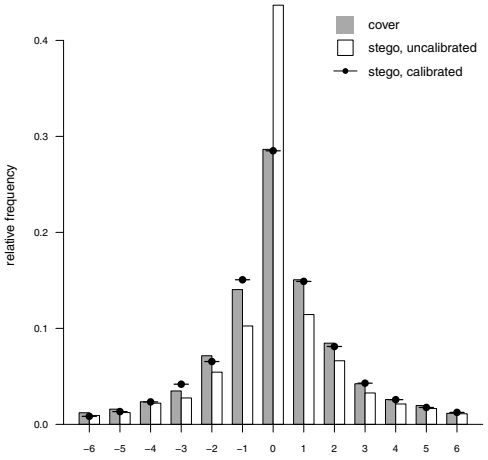
Up to now, contemporary techniques for digital steganography have been surveyed quite comprehensively. The remainder of this chapter is devoted to a description of the state of the art in steganalysis. This section introduces three basic techniques that have been developed specifically for the construction of steganalysis methods. Later, in Sect. 2.10, we present in greater detail a number of targeted detectors for LSB replacement steganography which are relevant to Part II of this book.

2.9.1 Calibration of JPEG Histograms

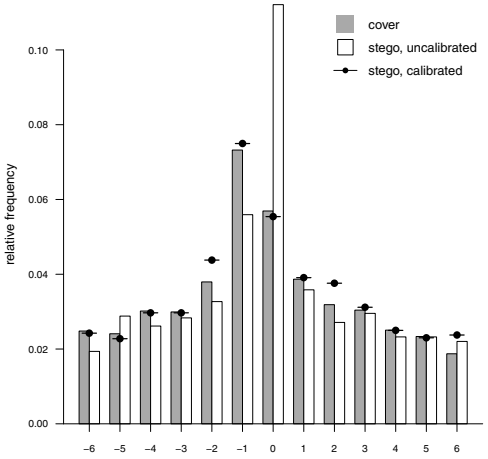
Calibration of JPEG histograms is a technique specific to steganalysis that was first introduced by Fridrich, Goljan, and Hogea [78] in their targeted detector against the F5 algorithm. It soon became a standard building block for many subsequent detectors against JPEG steganography, and is probably not limited to the JPEG domain, although applications in other transformed domains are rare due to the dominance of JPEG as a cover format in steganalysis research.

The idea of *calibration* is to estimate marginal statistics (histograms, co-occurrence matrices) of the *cover*'s transformed domain coefficients from the *stego* object by desynchronising the block transform structure in the spatial domain. The procedure works as depicted in Fig. 2.17. The suspected stego object in transformed domain representation is transferred back to the spatial domain (in the case of JPEG, a standard decompression operation), and then the resulting spatial domain representation is cropped by a small number

[37] The case for public-key steganography is less well understood, as pointed out in [16].



(a) AC subband $(3, 1)$



(b) AC subband $(1, 2)$

Fig. 2.16: Histograms of selected DCT subbands for a single JPEG image ($q = 0.8$). Its stego version is made by the F5 embedding operation ($p = 1$)

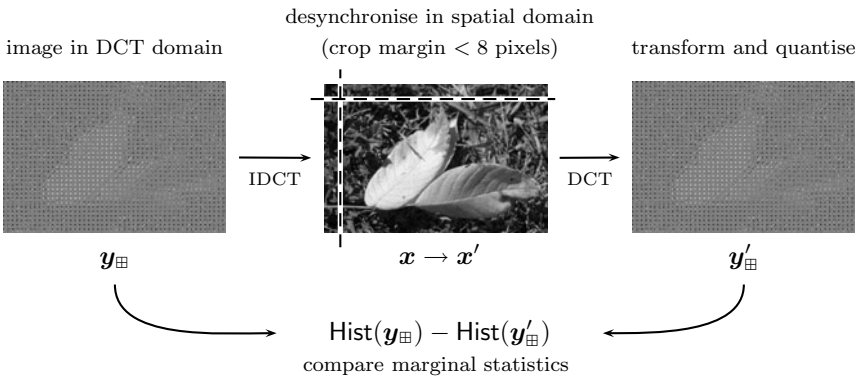


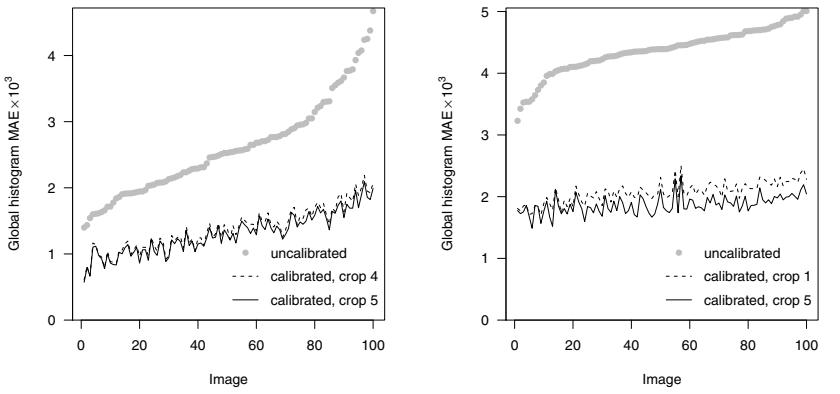
Fig. 2.17: Diagram of calibration procedure to estimate cover statistics

of pixels at two orthogonal margins. This ensures that the original (8×8) grid is desynchronised in a subsequent transformation to the transformed domain (re-compression for JPEG, using the same quantisation matrix as before). After this sequence of operations, the coefficients exhibit marginal statistics that are much closer to the original than those of the (suspected) stego object, where the repeated application of the embedding operation might have deformed the marginal statistics.

The capability of calibration to recover original histograms is shown in Fig. 2.16 (a) for selected subbands. As expected, the stego histogram is much more leptokurtic (the frequency of 0s increases) than the cover, which is a result of the moderated-sign embedding operation of the F5 algorithm used to produce the curves (cf. Fig. 2.13 (e), p. 44). The calibration procedure recovers the original values very accurately, so evaluating the difference between uncalibrated and calibrated histograms constitutes a (crude) detector.

Interestingly, the estimation is still acceptable—albeit not perfect—for 'abnormal' (more precisely, nonzero mode) histograms, as shown in Fig. 2.16 (b). A summary measure of the calibration procedure's performance can be computed from the global histogram mean absolute error (MAE) by aggregating the discrepancy between cover and stego estimates of all 63 AC DCT subbands. Quantitative results for a set of 100 randomly selected images are reported in Fig. 2.18 for different compression qualities and margin widths. Calibrated versions of the stego objects were evaluated for crop margins between one and six pixels. The curves show the margins that led to the best (solid line) and worst (dashed line) results. Tinkering with the margin width seems to yield small but systematic improvements for high compression qualities.

These and other experimental results confirm the effectiveness and robustness of calibrating JPEG histograms, but we are not aware of a rigourous



(a) $q = 0.8$ (b) $q = 0.95$

Fig. 2.18: Mean absolute error between normalised global AC DCT coefficient histogram of 100 JPEG cover images and simulated F5 stego objects ($p = 1$) with and without calibration for two different JPEG qualities q. Images are sorted by increasing uncalibrated MAE

mathematical analysis of the way calibration works. Known limitations of calibration include double-compressed JPEG images (with different quantisation matrices) and images that contain *spatial resonance*. This occurs when the content has a periodicity close to (an integer multiple of) the block size of the transformation. These phenomena as well as possible remedies are discussed in [77].

2.9.2 Universal Detectors

Steganalysis methods can be broadly divided into *targeted detectors*, which are designed to evaluate artefacts of particular embedding operations, and *universal detectors*, which do not assume prior knowledge about a particular steganographic system. Without a specific embedding operation to reverse engineer, universal methods extract from suspected stego objects a broad set of general statistical measures (so-called *features* $\boldsymbol{f} = (f_1, \ldots, f_k)$), which likely change after embedding. Often, features from different domains (spatial, various transforms) are combined in a feature vector. Then, a classifier

is trained with features from a large number of typical cover objects,[38] and classes are defined to distinguish between clean covers and stego objects. Training a classifier with a representative set of image data yields parameters $\boldsymbol{\theta}$, which are then used in a second stage to assign unknown objects to classes (cover or stego objects) according to their features. Proposals for the construction of classifiers are abundant in the machine learning literature. The most important types of classifiers employed in steganalysis include

- *ordinary least-squares regression* (OLS) and its refinement for classification purposes as *Fisher linear discriminant analysis* (FLD) [59], *quadratic discriminant analysis* (QDA) [201] and generalisations to *support vector machines* (SVM) [32] for continuous features,
- *Bayesian belief networks* (BBNs) [182] for discrete or discretised features, and
- *naïve Bayes classifiers* (NBCs) [49] for mixed feature vectors.

Researchers in the area of steganalysis have combined these machine learning techniques with a variety of features extracted from different domains of images and audio files [179].

Although suffering from lower detection reliability than decent targeted detectors, universal detectors have the advantage of easy adaptability to new embedding functions. While in this case targeted detectors have to be altered or redesigned, universal detectors just require a new training. Some critics argue that universal detectors are merely a combination of features known from published targeted detectors and hence are not as 'blind' as claimed.[39] So their ability to detect fundamentally new classes of embedding functions might be limited. Although there are few breakthroughs in the development of new embedding operations, experience with the introduction of new embedding domains, such as the MP domain proposed by Cancelli et al. [36], has shown that universal detectors that did not anticipate these innovations were not able to detect this new kind of steganography reliably (see also [191] for the difficulty of detecting 'minus-F5').

Table 2.2 (p. 62) summarises a literature review of the most relevant feature sets proposed for universal detectors of image steganography in the past couple of years. Note that we omit judgements about their performance as the authors did not use comparable image sets, embedding parameters, or evaluation procedures (e.g., testing different embedding functions independently

[38] The training objects comprise both clean covers and stego objects generated at the design stage of the method for training purposes. This implies that developers of universal detectors typically have access to actual steganographic systems or know their embedding operations.

[39] The name *blind detector* is used synonymously for universal detectors in the literature. We prefer the term 'universal' as rarely any detector in the literature has been designed without knowledge of the (set of) target embedding operations. What is more, in digital watermarking and multimedia forensics, the term 'blind' is reserved for detectors that work without knowledge of the original cover. In this sense, targeted detectors are also blind by definition.

Table 2.2: Overview of selected universal detectors for image steganography

Ref.	Method			Evaluation	
	feature description	classifier	# features	# images	tested stego systems
Avcibaş et al. [7]					
	spatial domain and spectral quality metrics	OLS	26	20	three watermarking algorithms
Lyu and Farid [163]					
	moments of DFT subband coefficients and size of predictor error	FLD, SVM	72	1,800	LSB, EzStego, JSteg, OutGuess
Harmsen and Pearlman [97]					
	HCF centre of mass (COM)	NBC	3	24	±1, SSIS, additive noise in DCT domain for RGB images
Chen et al. [40]					
	DCT moments, HCF moments, DWT HCF moments of image and prediction residual	SVM	260	798	LSB, ±1, SSIS, QIM, OutGuess, F5, MB1
Fridrich [68]					
	Delta to calibrated versions of DCT histogram measures, blockiness, coefficient co-occurrence	FLD	23	1,814	OutGuess, F5, MB1, MB2
Goljan et al. [91]					
	higher-order moments of residual from wavelet denoising filter	FLD	27	2,375	±1 and variants (side information, ternary codes, adaptive)
Shi et al. [215]					
	intra-block difference histograms of absolute DCT coefficients	SVM	324	7,560	OutGuess, F5, MB1
Pevný and Fridrich [187]					
	combination of [68] and [215]	SVM	274	3,400	Jphide, Steghide, F5, OutGuess, MB1, MB2
Lyu and Farid [164]					
	[163] plus LAHD phase statistics	SVM	432	40,000	JSteg, F5, Jphide, Steghide, OutGuess
Barbier et al. [10]					
	moments of residual entropy in Huffman-encoded blocks, KLD to reference p.d.f.	FLD	7+	4,000	F5, Jphide, OutGuess

or jointly). Another problem is the risk of overfitting when the number of images in the training and test set is small compared to the number of features, and all images are taken from a single source. In these cases, the parameters of the trained classifier are estimated with high standard errors and may be adapted too much to the characteristics of the test images so that the results do not generalise.

Although machine learning techniques were first used in steganalysis to construct universal detectors, they become increasingly common as tools for constructing targeted detectors as well. This is largely for convenience reasons: if several metrics sensitive to embedding are identified, but their optimal combination is unknown, then machine learning techniques help to find good decision rules quickly (though they are sometimes hard to explain). The ±1 detector proposed by Boncelet and Marvel [28] and the targeted detector of MB2 by Ullerich [227] are representatives of this approach.

The research in this book is restricted to targeted detectors, mainly because they have better performance than universal detectors and their higher transparency facilitates reasoning about dependencies between cover properties and detection performance.

2.9.3 Quantitative Steganalysis

The attribute *quantitative* in steganalysis means that the detector outputs not only a binary decision, but an estimate of the lengths of the secret message, which can be zero for clean covers [79]. This implies that those methods are still reliable when only parts of the cover's steganographic capacity have been used (early statistical detectors could only detect reliably messages with full capacity or imperfect spreading [238]).

We define quantitative detectors as functions that estimate the net embedding rate p. The attribute 'net' means that possible gains in embedding efficiency due to message coding (see Sect. 2.8.2) are not taken into account,

$$\hat{p} = \mathsf{Detect}_{\mathsf{Quant}}(\boldsymbol{x}^{(p)}). \tag{2.24}$$

A useful property of quantitative detectors is that detection performance can be measured more granularly than mere error rates, e.g., by comparing the estimated embedding rate p with the estimate $\hat{p}$. Quantitative detectors for a particular embedding operation, namely LSB replacement, play an important role in the specific results presented in Part II. Therefore, we introduce three state-of-the-art detectors and some variants in the next section.

2.10 Selected Estimators for LSB Replacement in Spatial Domain Images

We follow the terminology of Ker [120] and call a quantitative detector *estimator* when we refer to its ability to determine the secret message length, and *discriminator* when we focus on separating stego from cover objects.

2.10.1 RS Analysis

RS analysis,[40], developed by Fridrich, Goljan, and Du [74], estimates the number of embedding changes by measuring the proportion of *regular* and *singular* non-overlapping k-tuples (groups) of spatially adjacent pixels before and after applying three types of flipping operations:

1. $\mathsf{Flip}^{+1} : \mathcal{X} \to \mathcal{X}$ is a bijective mapping between pairs of values that mimics exactly the embedding operation of LSB replacement: $0 \leftrightarrow 1, 2 \leftrightarrow 3, \ldots$
2. $\mathsf{Flip}^{-1} : \mathcal{X} \to \mathcal{X}$ is a bijective mapping between the opposite (shifted) pairs, that is, $\mathsf{Flip}^{-1}(x) = \mathsf{Flip}^{+1}(x+1) - 1$; hence, $-1 \leftrightarrow 0, 1 \leftrightarrow 2, \ldots$
3. $\mathsf{Flip}^{0} : \mathcal{X} \to \mathcal{X}$ is the identity function.

Groups are counted as *regular* and assigned to multi-set $\mathcal{R}_{\mathbf{m}}$ if the value of a discrimination function $\mathsf{Discr} : \mathcal{X}^k \to \mathbb{R}$ increases after applying $\mathsf{Flip}^{\mathbf{m}_i}$ on the individual pixels of the group according to a mask vector $\mathbf{m} \in \{0,1\}^k$, i.e.,

$$\mathsf{Discr}\left(\boldsymbol{x}\right) < \mathsf{Discr}\left(\mathsf{Flip}^{\mathbf{m}_1}(x_1), \mathsf{Flip}^{\mathbf{m}_2}(x_2), \ldots, \mathsf{Flip}^{\mathbf{m}_k}(x_k)\right). \tag{2.25}$$

Conversely, multi-set $\mathcal{S}_{\mathbf{m}}$ contains all so-called *singular* groups, by definition, when

$$\mathsf{Discr}\left(\boldsymbol{x}\right) > \mathsf{Discr}\left(\mathsf{Flip}^{\mathbf{m}_1}(x_1), \mathsf{Flip}^{\mathbf{m}_2}(x_2), \ldots, \mathsf{Flip}^{\mathbf{m}_k}(x_k)\right). \tag{2.26}$$

The remaining *unusable* groups, for which none of inequalities (2.25) and (2.26) hold, is disregarded in the further analysis. The suggested implementation for the discrimination function is a noisiness measure based on the L_1-norm, but other summary functions are possible as well:

$$\mathsf{Discr}(\boldsymbol{u}) = \sum_{i=2}^{|\boldsymbol{u}|} |u_i - u_{i-1}|. \tag{2.27}$$

Figure 2.19 shows the typical shape of the relative sizes of $\mathcal{R}_{\mathbf{m}}$ (solid black curve) and $\mathcal{S}_{\mathbf{m}}$ (solid grey curve) as a function of the fraction of flipped LSBs

[40] RS stands for *regular/singular* named after the concept of regular and singular groups of pixels.

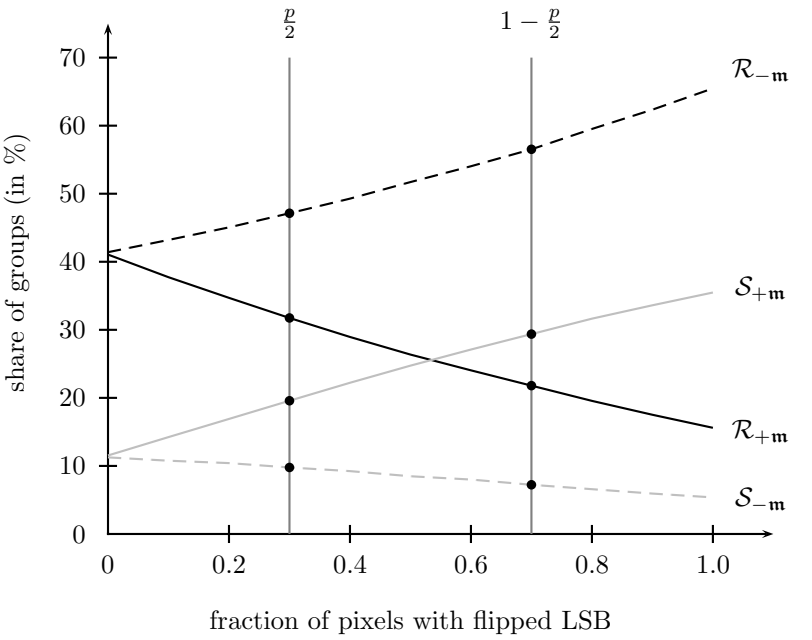


Fig. 2.19: Typical RS diagram of a single image: relative size of sets of regular ($\mathcal{R}$) and singular ($\mathcal{S}$) groups for direct ($+\mathbf{m}$) and inverse ($-\mathbf{m}$) mask $\mathbf{m} = (0, 1, 1, 0)$ as a function of the fraction of flipped LSBs

for a single image with non-overlapping horizontal groups of size $k = 4$ and mask $\mathbf{m} = (0, 1, 1, 0)$. The corresponding dashed curves $\mathcal{R}_{-\mathbf{m}}$ and $\mathcal{S}_{-\mathbf{m}}$ result from applying the *inverse mask* $-\mathbf{m} = (0, -1, -1, 0)$. LSB replacement is detectable because the proportion of regular and singular groups deviates in the opposite direction with increasing number of flipped LSBs.

The unknown embedding rate p of a suspect image $\boldsymbol{x}^{(p)}$ can be estimated from observable quantities in this diagram, a linear approximation of the 'outer' $\mathcal{R}_{-\mathbf{m}}$ and $\mathcal{S}_{-\mathbf{m}}$ curves as well as a quadratic approximation of the 'inner' curves $\mathcal{R}_{+\mathbf{m}}$ and $\mathcal{S}_{+\mathbf{m}}$.[41] The net embedding rate $\hat{p}$ is approximately half of the fraction of pixels with flipped LSBs.[42]

- The size of $\mathcal{R}_{+\mathbf{m}}$, $\mathcal{R}_{-\mathbf{m}}$, $\mathcal{S}_{+\mathbf{m}}$ and $\mathcal{S}_{-\mathbf{m}}$ at the intersection with the vertical line $p/2$ can be obtained directly from $\boldsymbol{x}^{(p)}$.

[41] The linear and quadratic shapes of the curves has been proven for groups of size $k = 2$ in [50]. More theory on the relation between the degree of the polynomial and the group size k is outlined in the introduction of [120].

[42] Net embedding rate and secret message length as a fraction of cover size n differ if efficiency-enhancing coding is employed; see Sect. 2.9.3.

- Flipping the LSBs of *all* samples in $\boldsymbol{x}^{(p)}$ and the subsequent calculation of multi-set sizes yield an indirect measure of the sizes of $\mathcal{R}_{+\mathbf{m}}$, $\mathcal{R}_{-\mathbf{m}}$, $\mathcal{S}_{+\mathbf{m}}$ and $\mathcal{S}_{-\mathbf{m}}$ at the intersection with the vertical line $1 - p/2$.

Further, two assumptions,

1. the two pairs of curves $\mathcal{R}_{\pm\mathbf{m}}$ and $\mathcal{S}_{\pm\mathbf{m}}$ intersect at 0 (a plausible assumption if we reckon that the distribution of intensity values in the image acquisition process is invariant to small additive constants), and
2. curves $\mathcal{R}_{+\mathbf{m}}$ and $\mathcal{S}_{+\mathbf{m}}$ intersect at 50% flipped LSBs (justified in [74] and [79] with a theorem cited from [90] saying that "the lossless capacity in the LSBs of a fully embedded image is zero"; in practice, this assumption is violated more frequently than the first one),

are sufficient to find a unique[43] solution for $\hat{p} = \frac{z}{z-1/2}$.

Auxiliary variable z is the smaller root of the quadratic equation

$$2(\Delta_{+\mathbf{m}} + \Delta'_{+\mathbf{m}})z^2 + (\Delta'_{-\mathbf{m}} - \Delta_{-\mathbf{m}} - \Delta_{+\mathbf{m}} - 3\Delta'_{+\mathbf{m}})z - \Delta'_{-\mathbf{m}} + \Delta'_{+\mathbf{m}} = 0 \quad (2.28)$$

$$\text{with} \qquad \Delta_{\mathbf{m}} = \frac{k}{n} \cdot (|\mathcal{R}_{\mathbf{m}}| - |\mathcal{S}_{\mathbf{m}}|) \text{ at } \frac{p}{2} \qquad \text{(computed from } \boldsymbol{x}^{(p)}\text{), and}$$

$$\Delta'_{\mathbf{m}} = \frac{k}{n} \cdot (|\mathcal{R}_{\mathbf{m}}| - |\mathcal{S}_{\mathbf{m}}|) \text{ at } 1 - \frac{p}{2} \qquad \text{(computed from } \mathsf{Flip}^{+1}(\boldsymbol{x}^{(p)})\text{).}$$

For p close to 1, cases where Eq. (2.28) has no real root occur more frequently. In such cases we set $\hat{p} = 1$ because the suspect image is almost certainly a stego image. However, failures of the RS estimation equation have to be borne in mind when evaluating the distribution of RS estimates and estimation errors $\hat{p} - p$, as done in Chapter 5.

The way pixels are grouped (topology and overlap), group size k, mask vector $\mathbf{m}$ and the choice of the discrimination function Discr (Eq. 2.27) are subject to experimental fine tuning. Empirical results can be found in [118] and [119]. Note that global RS estimates are not reliable if the message is not distributed randomly in the stego image. In this case a moving window variant of RS or SPA, as suggested in [79], or more efficient sequential variants of WS analysis [128, 133] are preferable.

[43] Yet another set of quantities could be obtained for 50% flipped LSBs by averaging over repeated randomisations of the entire LSB plane. Incorporating this information leads to an over-specified equation system for which a least-squares solution can be found to increase the robustness against measurement errors of individual quantities. Alternatively, the zero-intersection assumption can be weakened. Although there is little documented evidence on whether the performance gains justify the additional effort, the dominant variant of RS follows the approach described above. Research on RS improvements has stalled since more reliable detectors for LSB replacement have been invented.

2.10.2 Sample Pair Analysis

The steganalysis method known as *sample pair analysis*[44] (SPA) was first introduced by Dumitrescu et al. [50, 51]. In our presentation of the method we adapt the more extensible alternative notation of Ker [120] to our conventions.[45]

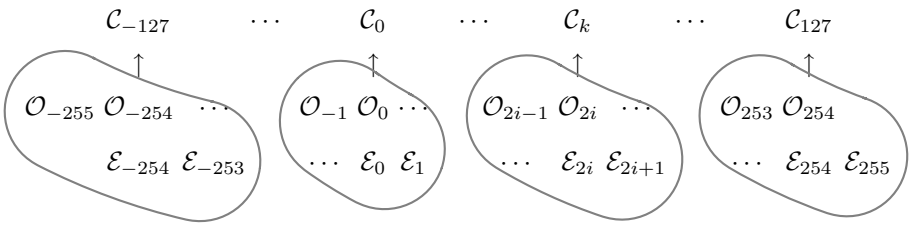


Fig. 2.20: Relation of trace sets and subsets in SPA ($\mathcal{X} = [0, 255]$)

Similarly to RS analysis, SPA evaluates groups of spatially adjacent pixels. It assigns each pair (x_1, x_2) to a *trace set* $\mathcal{C}_i$, so that

$$\mathcal{C}_i = \left\{ (x_1, x_2) \in \mathcal{X}^2 \;\middle|\; \left\lfloor \frac{x_2}{2} \right\rfloor - \left\lfloor \frac{x_1}{2} \right\rfloor = i \right\}, \quad |i| \leq \lfloor (\max \mathcal{X} - \min \mathcal{X})/2 \rfloor. \tag{2.29}$$

Each trace set $\mathcal{C}_i$ can be further partitioned into up to four *trace subsets*, of which two types can be distinguished:

- Pairs (x_1, x_2) whose values differ by $i = x_2 - x_1$ and whose first elements x_1 are *even* belong to $\mathcal{E}_i$.
- Pairs (x_1, x_2) whose values differ by $i = x_2 - x_1$ and whose first elements x_1 are *odd* belong to $\mathcal{O}_i$.

Consequently, the union of trace subsets $\mathcal{E}_{2i+1} \cup \mathcal{E}_{2i} \cup \mathcal{O}_{2i} \cup \mathcal{O}_{2i-1} = \mathcal{C}_i$ constitutes a trace set (cf. Fig. 2.20). This definition of trace sets and subsets ensures that the LSB replacement embedding operation never changes a sample pair's trace set, i.e., $\mathcal{C}_i^{(0)} = \mathcal{C}_i^{(p)} = \mathcal{C}_i$, but may move sample pairs between trace subsets that constitute the same trace set. So cardinalities $|\mathcal{C}_i|$ are invariant to LSB replacement, whereas $|\mathcal{E}_i|$ and $|\mathcal{O}_i|$ are sensitive. The transition probabilities between trace subsets depend on the net embedding rate p as depicted in the transition diagram of Fig. 2.21. So, the effect of

[44] The same method is sometimes also referred to as *couples analysis* in the literature to avoid possible confusion with *pairs analysis* by Fridrich et al. [82], another method not relevant in this book. Therefore, we stick to the original name.

[45] This presentation minds the order of samples in each pair; hence, i can be negative. The original publication made no difference between pairs (u, v) and (v, u). This led to a special case for $\lfloor u/2 \rfloor = \lfloor v/2 \rfloor$.

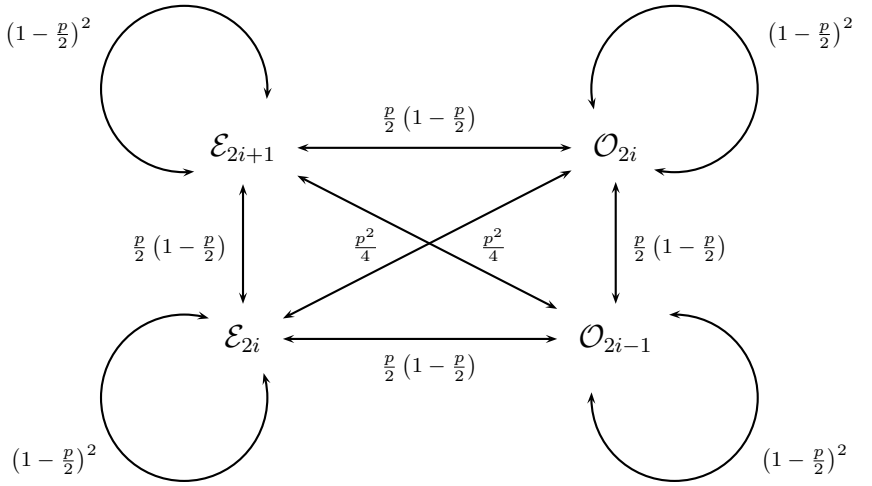


Fig. 2.21: Transition diagram between trace subsets under LSB replacement

applying LSB replacement with rate p on the expected cardinalities of the trace subsets can be written as four quadratic equations (in matrix notation):

$$\begin{bmatrix} |\mathcal{E}^{(p)}_{2i+1}| \\ |\mathcal{E}^{(p)}_{2i}| \\ |\mathcal{O}^{(p)}_{2i}| \\ |\mathcal{O}^{(p)}_{2i-1}| \end{bmatrix} = \begin{bmatrix} \left(1-\frac{p}{2}\right)^2 & \frac{p}{2}\left(1-\frac{p}{2}\right) & \frac{p}{2}\left(1-\frac{p}{2}\right) & \frac{p^2}{4} \\ \frac{p}{2}\left(1-\frac{p}{2}\right) & \left(1-\frac{p}{2}\right)^2 & \frac{p^2}{4} & \frac{p}{2}\left(1-\frac{p}{2}\right) \\ \frac{p}{2}\left(1-\frac{p}{2}\right) & \frac{p^2}{4} & \left(1-\frac{p}{2}\right)^2 & \frac{p}{2}\left(1-\frac{p}{2}\right) \\ \frac{p^2}{4} & \frac{p}{2}\left(1-\frac{p}{2}\right) & \frac{p}{2}\left(1-\frac{p}{2}\right) & \left(1-\frac{p}{2}\right)^2 \end{bmatrix} \begin{bmatrix} |\mathcal{E}^{(0)}_{2i+1}| \\ |\mathcal{E}^{(0)}_{2i}| \\ |\mathcal{O}^{(0)}_{2i}| \\ |\mathcal{O}^{(0)}_{2i-1}| \end{bmatrix}. \tag{2.30}$$

Trace subsets $\mathcal{E}^{(p)}$ and $\mathcal{O}^{(p)}$ are observable from a given stego object. An approximation of the cardinalities of the cover trace subsets $\mathcal{E}^{(0)}$ and $\mathcal{O}^{(0)}$ can be rearranged as a function of p by inverting Eq. (2.30). The transition matrix is invertible for $p < 1$:

$$\begin{bmatrix} |\hat{\mathcal{E}}^{(0)}_{2i+1}| \\ |\hat{\mathcal{E}}^{(0)}_{2i}| \\ |\hat{\mathcal{O}}^{(0)}_{2i}| \\ |\hat{\mathcal{O}}^{(0)}_{2i-1}| \end{bmatrix} = \frac{1}{(2-2p)^2} \begin{bmatrix} (2-p)^2 & p(p-2) & p(p-2) & p^2 \\ p(p-2) & (2-p)^2 & p^2 & p(p-2) \\ p(p-2) & p^2 & (2-p)^2 & p(p-2) \\ p^2 & p(p-2) & p(p-2) & (2-p)^2 \end{bmatrix} \begin{bmatrix} |\mathcal{E}^{(p)}_{2i+1}| \\ |\mathcal{E}^{(p)}_{2i}| \\ |\mathcal{O}^{(p)}_{2i}| \\ |\mathcal{O}^{(p)}_{2i-1}| \end{bmatrix}. \tag{2.31}$$

With one additional cover assumption, namely $|\mathcal{E}^{(0)}_{2i+1}| \approx |\mathcal{O}^{(0)}_{2i+1}|$, the first equation of this system for i can be combined with the fourth equation for $i+1$ to obtain a quadratic estimator $\hat{p}$ for p. This assumption mirrors the first assumption of RS analysis (see p. 66). It is plausible because cardinalities of

sample pairs in natural images should not depend on the parity of their first element:

$$|\hat{\mathcal{E}}^{(0)}_{2i+1}| = |\hat{\mathcal{O}}^{(0)}_{2i+1}| \tag{2.32}$$

$$\begin{aligned} 0 = {} & \frac{(2-p)^2}{(2-2p)^2}\left(|\mathcal{E}^{(p)}_{2i+1}| - |\mathcal{O}^{(p)}_{2i+1}|\right) + \frac{p^2}{(2-2p)^2}\left(|\mathcal{O}^{(p)}_{2i-1}| - |\mathcal{E}^{(p)}_{2i+3}|\right) + \\ & \frac{p(p-2)}{(2-2p)^2}\left(|\mathcal{E}^{(p)}_{2i}| + |\mathcal{O}^{(p)}_{2i}| - |\mathcal{E}^{(p)}_{2i+2}| - |\mathcal{O}^{(p)}_{2i+2}|\right) \end{aligned} \tag{2.33}$$

$$\begin{aligned} 0 = {} & p^2\left(|\mathcal{C}_i| - |\mathcal{C}_{i+1}|\right) + 4\left(|\mathcal{E}^{(p)}_{2i+1}| - |\mathcal{O}^{(p)}_{2i+1}|\right) + \\ & 2p\left(|\mathcal{E}^{(p)}_{2i+2}| + |\mathcal{O}^{(p)}_{2i+2}| - 2|\mathcal{E}^{(p)}_{2i+1}| + 2|\mathcal{O}^{(p)}_{2i+1}| - |\mathcal{E}^{(p)}_{2i}| - |\mathcal{O}^{(p)}_{2i}|\right) \end{aligned} \tag{2.34}$$

The smaller root of Eq. (2.34) is a secret message length estimate $\hat{p}_i$ based on the information of pairs in trace set $\mathcal{C}_i$. Standard SPA sums up the family of estimation equations (2.34) for a fixed interval around $\mathcal{C}_0$, such as $-30 \leq i \leq 30$, and calculates a single root $\hat{p}$ from the aggregated quadratic coefficients. Experimental results from fairly general test images have shown that standard SPA, using all overlapping horizontal and vertical pairs of greyscale images, is slightly more accurate than standard RS analysis [22, 118]. For solely discrimination purposes (hence, ignoring the quantitative capability), it has been found that smarter combinations of individual roots for small $|i|$, e.g., $\hat{p}^* = \min(\hat{p}_{-2}, \ldots, \hat{p}_2)$, can improve SPA's detection performance further [118].

Similarly to RS, Eq. (2.34) may fail to produce real roots, which happens more frequently as p approaches 1. In these cases, the tested object is almost certainly a stego image, but the exact message length cannot be determined.

2.10.3 Higher-Order Structural Steganalysis

Sample pair analysis, as presented in Sect. 2.10.2, is a specific representative of a family of detectors for LSB replacement which belong to the general framework of *structural steganalysis*. The attribute 'structural' refers to the design of detectors to deliberately exploit, at least in theory, all combinatorial measures of the artificial dependence between sample differences and the parity structure that is typical for LSB replacement.[46] A common element in all structural detectors is to estimate $\hat{p}$ so that macroscopic cover properties,

[46] Under LSB replacement (see Eq. 2.8), even cover samples are never decremented whereas odd cover samples are never incremented. This leads to the artificial parity structure.

which can be approximated from the stego object by inverting the effects of embedding as a function of p, match cover assumptions best. Hence, also RS analysis and the method by Zhang and Ping [252] (disregarded in this book) can be subsumed as (less canonical) representatives of the structural framework.[47] In this section we review three important alternative detectors of the structural framework, which are all presented as extensions to SPA.

2.10.3.1 Least-Squares Solutions to SPA

The combination of individual equations (2.34) for different i, as suggested in the original publication [51], appears a bit arbitrary. Lu et al. [160] have suggested an alternative way to impose the cover assumption $|\mathcal{E}_{2i+1}| \approx |\mathcal{O}_{2i+1}|$. Instead of setting both cardinalities equal, they argue that the difference between odd and even trace subsets should be interpreted as error,

$$\epsilon_i = |\mathcal{E}_{2i+1}| - |\mathcal{O}_{2i+1}|, \tag{2.35}$$

and a more robust estimate for $\hat{p}$ can be found by minimising the squared errors $\hat{p} = \arg\min_p \sum_i \epsilon_i^2$, which turns out to be a solution to a cubic equation. Note that the least-squares method (LSM) implicitly attaches a higher weight to larger trace subsets (those with small $|k|$ in natural images), where higher absolute deviations from the cover assumption are observable. Quantitative results reported in [160] confirm a higher detection accuracy in terms of MAE and estimator standard deviation than both RS and standard SPA for three image sets throughout all embedding rates p. In practice, pure LSM has shown to cause severe inaccuracies when p is close to 1, so a combination with standard SPA to screen for large embedding rates by a preliminary estimate is recommended in [22]. The combined method is called SPA/LSM.

2.10.3.2 Maximum-Likelihood Solutions to SPA

The process an image undergoes from acquisition via embedding to a stego object is indeterministic at many stages. The choice of the embedding positions and the encrypted message bits are (pseudo)random by definition to achieve secrecy. Additional parameters unknown to the steganalyst have to be modelled as random variables as well, foremost the cover realisation and the actual embedding rate p. A common simplification in the construction of

[47] At the time of this writing, it is unclear whether WS analysis (to be presented in the following section) belongs to the structural class (it probably does). WS was not well recognised when the structural terminology was introduced, so it is not commented on in [120]. Its different approach justifies it being treated as something special. However, variants of WS can be found that have a striking similarity to RS or SPA.

structural detectors is the (implicit) reduction of random variables to expectations. This is suboptimal as it ignores the shape of the random variables' probability functions, and their ad hoc algebraic combination may deviate from the true joint distribution. Moreover, deviations from the expectation are not weighted by the size of the standard error, which differs as trace sets are sparser populated for large $|i|$. As a remedy, Ker [126] has replaced the cover assumption $|\mathcal{E}_{2i+1}| = |\mathcal{O}_{2i+1}|$ by a probabilistic model in which all pairs in the union set $\mathcal{D}_{2i+1} = \mathcal{E}_{2i+1} \cup \mathcal{O}_{2i+1}$ are distributed uniformly into subsets $\mathcal{E}_{2i+1}$ and $\mathcal{O}_{2i+1}$ during an imaginary image acquisition process. The term 'pre-cover' has been suggested for the imaginary low-precision image composed of pairs in $\mathcal{D}_i$. With this model, probability functions for all random variables can be defined under gentle assumption and thus a likelihood function for structural detectors can be derived. Estimating $\hat{p}$ reduces to maximising the likelihood (ML).[48] As an additional advantage, likelihood ratio tests (LRTs) allow mathematically well-founded hypothesis tests for the existence of a stego message $p > 0$ against the null hypothesis $p = 0$ (though no practical tests exist that perform better than discriminators by the estimate $\hat{p}$, yet [126]).

Performance evaluations of a single implementation of SPA/ML suggest that ML estimates are much more accurate than other structural detectors, especially for low embedding rates p, where accuracy matters for discriminating stego images from plain covers. Unfortunately, the numerical complexity of ML estimates is high due to a large number of unknown parameters and the intractability of derivatives with respect to p. Computing a single SPA/ML estimate of a 1.0 megapixel image takes about 50 times longer than a standard SPA estimate [126]. However, more efficient estimation strategies using iteratively refined estimates for the unknown cardinalities $|\mathcal{D}_i|$ (e.g., via the expectation maximisation algorithm [47]) are largely unexplored and promise efficiency improvements in future ML-based methods. All in all, structural ML estimators are rather novel and leave open questions for research.

Earlier non-structural proposals for maximum-likelihood approaches to detect LSB replacement in the spatial domain [46, 48] work solely on the first and second order (joint) histograms and are less reliable than the ML-variant of SPA, which uses trace subsets to exploit the characteristic parity structure.

2.10.3.3 Triples and Quadruples Analysis

The class of structural detectors can be extended by generalising the principles of SPA from pairs to k-tuples [120, 122]. Hence, trace sets and subsets are indexed by $k-1$ suffixes and the membership rules generalise as follows:

[48] As argued in [126], the least-squares solution concurs with the ML estimate only in the case of independent Gaussian variables, but the covariance matrix contains nonzero elements for structural detectors.

$$\mathcal{C}_{i_1,\ldots,i_{k-1}} = \left\{(x_1,\ldots,x_k) \in \mathcal{X}^k \;\middle|\; \left\lfloor \frac{x_{j+1}}{2} \right\rfloor - \left\lfloor \frac{x_j}{2} \right\rfloor = i \;\forall j : 1 \le j < k \right\}$$
$$\mathcal{E}_{i_1,\ldots,i_{k-1}} = \left\{(x_1,\ldots,x_k) \in \mathcal{X}^k \;\middle|\; x_{j+1} - x_j = i \;\forall j : 1 \le j < k \;\wedge x_1 \text{ even} \right\}$$
$$\mathcal{O}_{i_1,\ldots,i_{k-1}} = \left\{(x_1,\ldots,x_k) \in \mathcal{X}^k \;\middle|\; x_{j+1} - x_j = i \;\forall j : 1 \le j < k \;\wedge x_1 \text{ odd} \right\}$$

Each trace set contains 2^k trace subsets. The generalisation of the transition matrix of Eq. (2.30) is given by the iterative rule $\boldsymbol{t}_k(p) = \boldsymbol{t}_{k-1}(p) \otimes \boldsymbol{t}_1(p)$ with initial condition

$$\boldsymbol{t}_1 = \begin{bmatrix} 1 - \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & 1 - \frac{p}{2} \end{bmatrix}. \tag{2.36}$$

For example, when $k = 3$, each trace set is divided into eight trace subsets with transition probabilities

- $\left(1 - \frac{p}{2}\right)^3$ for remaining in the same trace subset (no LSB flipped),
- $\frac{p}{2}\left(1 - \frac{p}{2}\right)^2$ for a move into a subset that corresponds to a single LSB flip,
- $\frac{p^2}{4}\left(1 - \frac{p}{2}\right)$ for a move into a subset where two out of three LSBs are flipped, and
- $\frac{p^3}{8}$ for a move to the 'opposite' trace subsets, i.e., with all LSBs flipped.

The corresponding transition diagram is depicted in Fig. 2.22. Selected transition paths are plotted and annotated only for trace subset $\mathcal{O}_{2i-1,2j}$ to keep the figure legible.

Inverting the transition matrix is easy following the procedure of [120]. A more difficult task for higher-order structural steganalysis is finding (all) equivalents for the cover assumption $|\mathcal{E}_{x_1,\ldots,x_{k-1}}| \approx |\mathcal{O}_{x_1,\ldots,x_{k-1}}|$. Apart from this *parity symmetry*, Ker [122] has identified two more classes of plausible cover assumptions, which he calls *inversion symmetry* and *permutative symmetry*. Once all relevant symmetries are identified, the respective estimation equations similar to Eq. (2.34) can be derived and solved either by ad hoc summation, the above-described least-squares fit, or through an ML estimate.

In general, higher-orders of structural steganalysis yield moderate performance increases, especially for low embedding rates, but for increasing k, their applicability reduces to even lower ranges of p. Another drawback of higher-orders is the low number of observations in each subset, which increasingly thwarts the use of the law of large numbers that frequencies converge towards their expected value, and the normal approximation for the multinomial distributions in the ML estimator. So, we conjecture that the optimal order k should depend on the size of the stego objects under analysis.

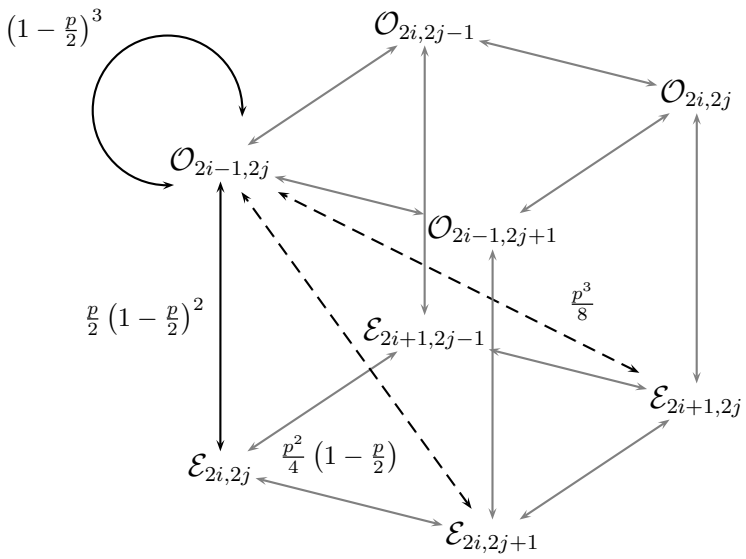


Fig. 2.22: Transition cube of trace subsets for Triples analysis ($k = 3$)

2.10.4 Weighted Stego Image Steganalysis

The steganalysis method using a weighted stego image (WS) proposed by Fridrich and Goljan [73] in 2004 differs from the above-discussed methods in several aspects: it is a mathematically better founded, modular, and computationally fast estimator for the net embedding rate of LSB replacement steganography in the spatial domain. In its original form, its performance is competitive with alternative methods only at high embedding rates, where high accuracy is less relevant in practice. Thus, the method resided in the shade for years. In this section we describe *standard WS* in an extensible notation. Improvements of the method are presented in Chapter 6.

WS analysis is based on the following concepts:

- A *weighted stego image* with scalar parameter λ:

$$\boldsymbol{x}^{(p,\lambda)} = \lambda\, \overline{\boldsymbol{x}}^{(p)} + (1-\lambda)\, \boldsymbol{x}^{(p)}, \tag{2.37}$$

  where $\overline{x} = x + (-1)^x = \mathsf{Flip}^{+1}(x)$, also applicable to vectors $\boldsymbol{x}$, is defined as a sample with inverted LSB to simplify the notation.
- Function $\mathsf{Pred} : \mathcal{X}^n \to \mathcal{X}^n$, a local predictor for pixels in cover images from their spatial neighbourhood.

- Function $\mathsf{Conf} : \mathcal{X}^n \to \mathbb{R}^{+n}$, a measure of local predictability with respect to Pred. By convention, lower values denote higher confidence or predictability.

The WS method is modular as Pred and Conf can be adapted to specific cover models while maintaining the same underlying logic of the estimator. Theorem 1 of [73] states the key idea of WS, namely that $\hat{p}$ can be estimated via the weight λ that minimises the Euclidean distance between the weighted stego image $\boldsymbol{x}^{(p,\lambda)}$ and the cover $\boldsymbol{x}^{(0)}$:

$$\hat{p} = 2 \arg\min_{\lambda} \sum_{i=1}^{n} \left(x^{(p,\lambda)} - x^{(0)} \right)^2 . \tag{2.38}$$

The proof of this theorem is repeated in Appendix C, using our notation. In practice, the steganalyst does not know the cover $\boldsymbol{x}^{(0)}$, so it has to be estimated from the stego object $\boldsymbol{x}^{(p)}$ itself. According to Theorem 3 in [73], the relation in Eq. (2.38) between $\hat{p}$ and λ still holds approximately if

1. $\boldsymbol{x}^{(0)}$ is replaced by its prediction $\mathsf{Pred}(\boldsymbol{x}^{(p)})$, and (independently)
2. the L_2-norm itself is weighted by vector $\boldsymbol{w}$ to reflect heterogeneity in predictability of individual samples.[49]

So, we obtain the main estimation equation that is common to all WS methods:

$$\begin{aligned} \hat{p} &= 2 \arg\min_{\lambda} \sum_{i=1}^{n} w_i \left(x_i^{(p,\lambda)} - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i \right)^2 && (2.39) \\ &= 2 \arg\min_{\lambda} \sum_{i=1}^{n} w_i \left(\lambda\, \overline{x}_i^{(p)} + (1-\lambda)\, x_i^{(p)} - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i \right)^2 \\ &= 2 \sum_{i=1}^{n} w_i \left(x_i^{(p)} - \overline{x}_i^{(p)} \right) \left(x_i^{(p)} - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i \right), && (2.40) \end{aligned}$$

where weights $\boldsymbol{w} = (w_1, \ldots, w_n)$ are calculated from the predictability measure as follows:

$$w_i \propto \frac{1}{1 + \mathsf{Conf}(\boldsymbol{x}^{(p)})_i} \quad , \text{ so that } \quad \sum_{i=1}^{n} w_i = 1. \tag{2.41}$$

In standard WS, function Pred is instantiated as the unweighted mean of the four directly adjacent pixels (in horizontal and vertical directions, ignoring diagonals). More formally,

[49] These optional local weights w_i should not be confused with the global weight λ that lends its name to the method. This is why the seemingly counterintuitive term 'unweighted weighted stego image steganalysis' makes sense: it refers to WS with constant local weight $w_i = 1/n \ \forall i$ (still using an estimation via λ).

$$\mathsf{Pred}(\boldsymbol{x}) = \boldsymbol{\Phi}\,\boldsymbol{x} \oslash \boldsymbol{\Phi}\,\mathbf{1}_{n\times 1}, \tag{2.42}$$

where $\boldsymbol{\Phi}$ is a $n \times n$ square matrix and $\Phi_{ij} = 1$ if the sample $\boldsymbol{x}_j^{(p)}$ is an upper, lower, left or right direct neighbour of sample $\boldsymbol{x}_i^{(p)}$; otherwise, $\Phi_{ij} = 0$. Operator $\oslash$ denotes element-wise division. Consistent with the choice of Pred, function Conf measures predictability as the empirical variance of all pixels in the local predictor; thus,

$$\mathsf{Conf}(\boldsymbol{x}) = \left(\frac{1}{n}\right)\left[\left((\boldsymbol{x}\otimes\mathbf{1}_{1\times n})\odot\boldsymbol{\Phi}\right)^2\mathbf{1}_{n\times 1}\right] - \left(\frac{1}{n^2}\right)\left[\left((\boldsymbol{x}\otimes\mathbf{1}_{1\times n})\odot\boldsymbol{\Phi}\right)\mathbf{1}_{n\times 1}\right]^2 \tag{2.43}$$

It is important to note that both the local prediction Pred and the local weights w_i must not depend on the value of $x_i^{(p)}$. Otherwise, correlation between the predictor error in covers $\mathsf{Pred}(\boldsymbol{x}^{(0)}) - \boldsymbol{x}^{(0)}$ and the parity of the stego sample $\boldsymbol{x}^{(p)} - \overline{\boldsymbol{x}}^{(p)}$ accumulates to a non-negligible error term in the estimation relation Eq. (2.40), which can be rewritten as follows to study the error components (cf. Eq. 6 of [73]):

$$\hat{p} = \overbrace{2\sum_{i=1}^{n} w_i\left(x_i^{(p)} - \overline{x}_i^{(p)}\right)\left(x_i^{(p)} - x_p^{(0)}\right)}^{\approx p} + \tag{2.44}$$
$$2\sum_{i=1}^{n} w_i\left(x_i^{(p)} - \overline{x}_i^{(p)}\right)\Big(\underbrace{x_p^{(0)} - \mathsf{Pred}(\boldsymbol{x}^{(0)})_i}_{\text{predictor error}} + \underbrace{\mathsf{Pred}(\boldsymbol{x}^{(0)})_i - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i}_{\text{predicted stego noise}}\Big).$$

Choosing functions Pred and Conf to be independent of the centre pixel bounds the term annotated as 'predictor error'. The term 'predicted stego noise' causes an estimation *bias* in images with large connected areas of constant pixel intensities,[50] for example, as a result of saturation. Imagine a cover where all pixels are constant and even, $x_i^{(0)} = 2k\ \forall i$ with k integer. With Pred as in Eq. (2.42), the prediction error in the cover $x_i^{(0)} - \mathsf{Pred}(\boldsymbol{x}^{(0)})_i = 0$, but the predicted stego noise $\mathsf{Pred}(\boldsymbol{x}^{(0)})_i - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i$ is negative on average because $\mathsf{Pred}(\boldsymbol{x}^{(0)})_i = 2k\ \forall i$ and $\mathsf{Pred}(\boldsymbol{x}^{(p)})_i = 2k$ with probability $(1 - p/2)^4$ (none of the four neighbours flipped), or $2k < \mathsf{Pred}(\boldsymbol{x}^{(p)})_i \leq 2k+1$ otherwise. With $w_i = 1/n\ \forall i$, the remaining error term,

$$\frac{2}{n}\sum_{i=1}^{n}\left(x_i^{(p)} - \overline{x}_i^{(p)}\right)\left(\mathsf{Pred}(\boldsymbol{x}^{(0)})_i - \mathsf{Pred}(\boldsymbol{x}^{(p)})_i\right) > 0 \quad \text{for } p > 0, \tag{2.45}$$

cancels out only for $p \in \{0, 1\}$. The size of the bias in real images depends on the proportion of flat areas relative to the total image size. Fridrich and

[50] Later, in Chapter 6, we argue that a more precise criterion than flat pixels is a phenomenon we call *parity co-occurrence*, which was not considered in the original publication.

Goljan [73] propose a heuristic bias correction, which estimates the number of flat pixels in $\boldsymbol{x}^{(0)}$ from the number of flat pixels in $\boldsymbol{x}^{(p)}$, although they acknowledge that their estimate is suboptimal as flat pixels can also appear randomly in $\boldsymbol{x}^{(p)}$ if the cover pixel is not flat. While this correction apparently removes outliers in the test images of [73], we could not reproduce improvements of estimation accuracy in our own experiments.

Compared to other quantitative detectors for LSB replacement, WS estimates are equally accurate even if the message bits are distributed unevenly over the cover. By adapting the form of Eq. (2.40) to the embedding hypothesis, WS can further be specialised to so-called *sequential embedding*, which means that the message bits are embedded with maximum density (i.e., change rate $1/2 \leftrightarrow p = 1$ in the local segment) in a connected part of the cover. This extension increases the detection accuracy dramatically (by about one order of magnitude), with linear running time still, even if both starting position and length of the message are unknown [128, 133]. Another extension to WS is a generalisation to mod-k replacement proposed in [247].

2.11 Summary and Further Steps

If there is one single conclusion to draw from this chapter, then it should be a remark on the huge design space for steganographic algorithms and steganalytic responses along possible combinations of cover types, domains, embedding operations, protocols, and coding. There is room for improvement in almost every direction. So, it is only economical to concentrate on understanding the building blocks separately before studying their interactions when they are combined. This has been done for embedding operations, and there is also research targeted to specific domains (MP [35, 36], YASS [218]) and coding (cf. Sect. 2.8.2). This book places an emphasis on covers because they are relevant and not extensively studied so far.

To study heterogeneous covers systematically, we take a two-step approach and start with theoretical considerations before we advance to practical matters. One problem of many existing theoretical and formal approaches is that their theorems are limited to artificial channels. In practice, however, high-capacity steganography in empirical covers is relevant. So, our next step in Chapter 3 is to reformulate existing theory so that it is applicable to empirical covers and takes account of the uncertainty.

The second step is an experimental validation of our theory: Chapters 4 to 7 document advances in statistical steganalysis. Our broader objective is to develop reusable methodologies, and provide proof of concepts, but we have no ambition to exhaustively accumulate facts. Similarly to the design space for steganographic algorithms, the space of possible aspects of heterogeneity in covers is vast. So closing *all* gaps is unrealistic—and impossible for empirical covers, as we will argue below.

Remark: Topics Excluded or Underrepresented in this Chapter

Although this chapter might appear as a fairly comprehensive and structured summary of the state of the art in steganography and steganalysis to 2009, we had to bias the selection of topics towards those which are relevant to the understanding of the remaining parts of this book. So we briefly name the intentionally omitted or underrepresented topics as a starting point for interested readers to consult further sources.[51]

We have disregarded the attempts to build provably secure steganography because they fit better into and depend on terminology of Chapter 3. Embedding operations derived from watermarking methods (e.g., the Scalar Costa scheme or quantisation index modulation) have been omitted. Robust steganography has not received the attention it deserves, but little is published for practical steganography. Research on the borderline between covert channels and digital steganography (e.g., hidden channels in games or network traffic [174]) is not in the scope of this survey. Finally, a number of not seriously tested proposals for adaptive or multi-sample embedding functions has probably missed our attention. Quite a few of such proposals were presented at various conferences with very broad scope: most of these embedding functions would barely be accepted at venues where the reviewers consider steganography a security technique, not a perceptual hiding exercise.

[51] We also want to point the reader to a comprehensive reference on modern steganography and steganalysis. The textbook by Jessica Fridrich [70] was published when the manuscript for this book was in its copy-editing phase.