Electronic Colloquium on Computational Complexity, Report No. 10 (2015)

Security Levels in Steganography – Insecurity does not Imply Detectability

Maciej Liśkiewicz, Rüdiger Reischuk, and Ulrich Wölfel

Institut für Theoretische Informatik, Universität zu Lübeck
Ratzeburger Allee 160, 23538 Lübeck, Germany
liskiewi@tcs.uni-luebeck.de, reischuk@tcs.uni-luebeck.de, uwoelfel@gmx.de

Abstract. This paper takes a fresh look at security notions for steganography – the art of encoding secret messages into unsuspicious covertexts such that an adversary cannot distinguish the resulting stegotexts from original covertexts. Stegosystems that fulfill the security notion used so far, however, are quite inefficient. This setting is not able to quantify the power of the adversary and thus leads to extremely high requirements. We will show that there exist stegosystems that are not secure with respect to the measure considered so far, still cannot be detected by the adversary in practice.
This indicates that a different notion of security is needed which we call *undetectability*. We propose different variants of (un)-detectability and discuss their appropriateness. By constructing concrete examples of stegosystems and covertext distributions it is shown that among these measures only one manages to clearly and correctly differentiate different levels of security when compared to an intuitive understanding in real life situations. We have termed this *detectability on average*.
As main technical contribution we design a framework for steganography that exploits the difficulty to learn the covertext distribution. This way, for the first time a tight analytical relationship between the task of discovering the use of stegosystems and the task of differentiating between possible covertext distributions is obtained.

Keywords: steganography, covertext channels, provable security, distribution learning.

1 Introduction

In cryptography the notion of *security* is well understood. Roughly speaking, a secure cryptosystem is defined by the property that an adversary with bounded resources cannot decipher a secret message. If a cryptosystem is not secure then there exists such an adversary with a significant advantage over random guessing. Considering a cryptosystem as a game between the encoder Alice and an adversary, this dichotomy looks natural: either there is an opponent with an advantage to decipher the secret message or not.

Security becomes a much more challenging property if one considers steganography, where secret messages are hidden into unsuspicious covertexts (see [4, 9, 1, 2, 8, 5, 11, 13] for theoretical foundations of steganography and some of the achievements of recent years). Here the adversary should not be able to distinguish between the resulting stegotexts and original covertexts that are exchanged between the stegoencoder Alice and the recipient Bob. Steganographic security crucially depends on properties of the covertext channel, also called the covertext distribution – a stegosystem might be much more secure for one channel than another, even if both look similar. If the covertext distribution is uniformly random over the channel alphabet, secure steganography becomes almost trivial. Thus, for this particular channel efficient and secure steganography is possible. However, in practice meaningful covertexts like natural pictures or speech do not allow arbitrary combinations of pixels or tones to make up a meaningful covertext. Requiring that a

ISSN 1433-8092

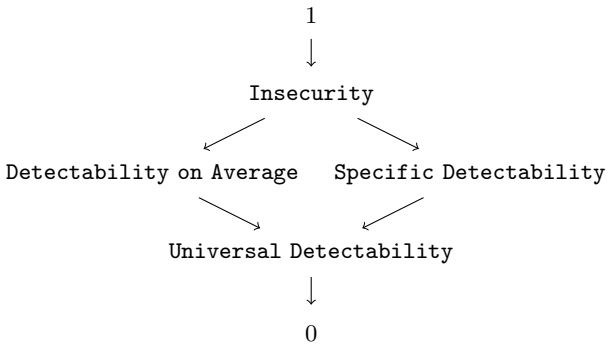
stegosystem should work for every channel is quite a strong demand. We will argue that under this condition secure steganography cannot be efficient. Therefore, it is important to analyse precisely the setting of the game between the stegoencoder and the adversary, and in particular to determine the level of influence that the stegoencoder has in choosing the covertext channel. In cryptography, to the contrary, the channel distribution is simply determined by the cryptosystem and the chosen key. By Kerckhoffs' principle it is assumed that the distribution is completely known to all parties.

We will show below that the use of a stegosystem that is *insecure* according to the definition used so far might still not be detected by an adversary. Up to now, a stegosystem is defined as insecure if the strongest possible adversary can detect the use of steganography. It suffices if this is true for a single channel chosen among all possible channels. However, there might be channels for which the adversary does not have a good chance for detection. It seems unrealistic that a stegoencoder would only make use of covertext channels that are easy to detect. This observation leads us to the question how an appropriate notion of security should look like and when to call a stegosystem insecure in practice.

Why do we see the need for considering *insecure* stegosystems (according to the current definition), although secure stegosystems have already been established (see e.g. Hopper et al. [9])? The answer is that security is only one of several desirable properties of a stegosystem. A "useful" stegosystem should also be *reliable* (i.e., with high probability, embedded messages can be reconstructed correctly), *efficient* (i.e., the time, space and oracle query complexities should be polynomial in the length of the hidden message) and achieve a good *transmission rate* for the hidden messages (i.e., the ratio between message bits per covertext and covertext entropy should not be too small).

Previously proposed stegosystems fail to meet all these criteria simultaneously. The formal model for stegosystems makes use of a *conditional sampling oracle*. Such an oracle receives as input a history $\mathcal{H}$ of previously drawn covertext documents and returns the next document based on this history. In order to maintain a good transmission rate, in each covertext document one should be able to embed an amount of information that grows with the document entropy. Since the stegosystem in [9] embeds only 1 bit of hidden information per document, Dedić et al. [5] have analysed the case where b bits have to be embedded per covertext in order to keep a good transmission rate. For a natural adaptation of the stegosystem in [9] they have obtained an exponential (in b) query complexity for the covertext oracle. Thus, the resulting stegosystem has a good rate, is secure and reliable, but very inefficient. In fact, they have shown that in the common black-box setting this exponential sampling complexity holds for *all secure* stegosystems. In this model the stegoencoder has no knowledge whatsoever about the covertext channel (except its min-entropy) and can only access it via a sampling oracle while the adversary is supposed to know everything about the channel. In particular, this leads to the strong conclusion that all schemes used in practice are insecure if security is defined based on this extreme setting.

Thus, due to the results in [5] there is no hope for efficient, practically usable steganography that can be proven secure, according to the definition used so far. But does this mean that in practice a steganalyst will be successful? In particular, can one conclude that every stegosystem, like e.g. the popular F5 scheme designed to hide information in JPEG images [15], is always breakable? In [12] we have discussed these questions and proposed a new model to resolve the highly unbalanced knowledge about the covertext channel by the adversary and the encoder. In the steganographic *grey-box model* the encoder starts with some partial knowledge at least about the type of covertext channel. In this paper we further elaborate on these problems and take the perspective of the



	Knowledge about channel	
	Alice	Adversary
`Insecurity`	–	×
`Detectability on Average`	–	–
`Specific Detectability`	×	×
`Universal Detectability`	×	–

Fig. 1. Relationship among different security levels for a stegosystem $\mathcal{S}$ and the state of knowledge about the covertext channel by the stegoencoder Alice and the adversary. The arrows indicate the relation between values of the corresponding (in)security measures for $\mathcal{S}$, where 1 and 0 denote the highest, resp. lowest values. The dashes and crosses in the table mean 'no knowledge', resp. 'full knowledge' about the covertext channel. According to the definition of `Insecurity` used so far, $\mathcal{S}$ is *insecure*, if there *exists* a single channel $\mathcal{C}_0$ such that the adversary using some specific strategy which may depend on $\mathcal{S}$ *and* on $\mathcal{C}_0$, can detect the stegosystem $\mathcal{S}$ over $\mathcal{C}_0$. However, this does *not* imply that the adversary can detect the usage of $\mathcal{S}$ for any other channel $\mathcal{C}$.

adversary in order to investigate how successful he can be in detecting steganography. This deficit has also been noticed in [10] trying to *move steganalysis from the laboratory into the real world.* The authors state that *almost all current steganalysis literature adheres to the model . . . so that the steganalyst can only learn about the cover source by empirical samples.*

To further develop a suitable formal model we introduce the concept of *detectability* and give three possible definitions *channel universal detectability*, *channel specific detectability* and *detectability on average* (see Fig. 1). They will be used in analysing the interplay between insecurity and detectability of stegosystems. Investigating possible definitions of detectability, we show how these measures relate to each other and that one of them, *detectability on average*, clearly outperforms the others. This measure also corresponds better to real life intuition of insecurity than the definition used so far and in fact, its assumptions are already used implicitly in applied steganalysis.

We construct an efficient stegosystem for the family of channels used by Dedić et al. for which the insecurity has to be large according to [5]. But we will show that its detectability on average is small, i.e., such a stegosystem can still be considered secure enough, as most of the time the adversary has no better chance than random guessing.

The paper is organised as follows. Basic notation and concepts of steganography are given in the next section. A review of common definitions of security in steganography and the introduction of our new measures for detectability is presented in Section 3. Then a tight relationship with the ability to distinguish different channels is established. In Section 5 we consider a natural family of so-called flat h-channels and stegosystems for this family to compare these measures. The analysis in this section is quite technical, and the reader should be familiar with the concepts introduced in [5]. Finally, in Section 7 we make some concluding remarks and indicate future research directions.

2 Basic Notation and Definitions

The definitions of the basic steganography concepts: channel, stegosystem, reliability, and insecurity of a stegosystem used in this paper are essentially those of [9] with modifications as proposed in [5].

Let Σ be a finite alphabet, Σ^ℓ (resp. $\Sigma^\star$) the set of strings of length ℓ (resp. finite length) over Σ, and $\sigma := \log|\Sigma|$. We denote the length of a string u by $|u|$ and the concatenation of two strings u_1 and u_2 by $u_1||u_2$, or by u_1u_2 if this does not lead to ambiguities. Symbols $u \in \Sigma$ will be called *documents* and a finite concatenation of documents a *communication sequence* or *covertext*. Typically, the document models a piece of data (e.g. a digital image or fragment of the image) while the communication sequence $c \in \Sigma^\star$ models the complete message sent to the receiver in a single communication exchange. If D is a probability distribution with finite support denoted by $\mathrm{supp}(D)$, we define the *min-entropy* $H_\pi(D)$ of D as the value $H_\pi(D) = \min_{x\in\mathrm{supp}(D)} - \log \Pr_D[x]$.

Definition 1 (Channel). *A channel $\mathcal{C}$ is a function that takes a history $\mathcal{H} \in \Sigma^\star$ as input and produces a probability distribution $D_\mathcal{H}$ on Σ. A history $\mathcal{H} = c_1c_2\ldots c_m$ is legal if each subsequent symbol is obtainable given the previous ones, i.e., $\Pr_{D_{c_1c_2\ldots c_{i-1}}}[c_i] > 0$ for all $i \le m$. The min-entropy of $\mathcal{C}$ is the value $\min_\mathcal{H} H_\pi(D_\mathcal{H})$ where the minimum is taken over all legal histories $\mathcal{H}$.*

This gives a very general definition of covertext distributions which allows dependencies between individual documents that are present in typical real-world communications.

Example 1. Let us assume our channel $\mathcal{C}$ describes valid, meaningful sentences in the English language (ignoring punctuation marks). The set of documents consists of all possible English words. Now, let the history $\mathcal{H}$ consist of the following beginning of a sentence: "I am standing on the". The distribution produced by $\mathcal{C}$ will probably give words like "grass", "peak" or "right" a high probability, as these words would likely be expected given $\mathcal{H}$. Less likely, but still with positive probability (because they are grammatically correct given $\mathcal{H}$) would be words like "needle", "door" or "justice". However, words like "an", "make" or "why" would be grammatically incorrect in the context of $\mathcal{H}$ and therefore associated with probability 0. Note, that the history $\mathcal{H}$ is legal since each subsequent word of $\mathcal{H}$ is obtainable given the previous ones, e.g. "I am" can be extended, with a positive probability, with word "standing".

In order to embed additional information into covertexts, one has to assume that the covertext channel distribution has a sufficiently large min-entropy. To get information about the covertext distribution *sampling oracles* can be used. $EX_\mathcal{C}(\mathcal{H})$ denotes an oracle that generates documents according to a channel $\mathcal{C}$ with history $\mathcal{H}$, i.e. each call of $EX_\mathcal{C}(\mathcal{H})$ returns a document c with probability $\Pr_{D_\mathcal{H}}[c]$ and the responses are independent of each other. A steganographic information transmission is thought of as taking a finite sequence $C_1, C_2, \ldots \in \Sigma^\star$ of covertexts and based on them to construct a stegotext $S \in \Sigma^\star$ such that the sequence additionally encodes an independent message M. This encoding is done by Alice who then sends the stegotext to the receiver Bob over a public channel. Let b denote the message encoding rate, i.e. a single stegodocument can encode up to b bits of M. Longer messages M have to be split into blocks of b bits each and for each block a separate stegodocument is generated. Their concatenation yields the stegotext.

Definition 2 (Stegosystem). *In the following, let $n = \ell \cdot b$ denote the length of the messages to be embedded, thus ℓ stegodocuments each hiding b bits are needed. A stegosystem $\mathcal{S}$ for the message space $\{0,1\}^n$ is a pair of probabilistic algorithms $[SE, SD]$ with the following functionality:*

- *$SE = SE(K, M, \mathcal{H})$ is the encoding algorithm that takes as input a randomly chosen secret key $K \in \{0,1\}^\kappa$ of length κ, where κ is a security parameter that depends on n, a message $M \in \{0,1\}^n$ (called hiddentext), a channel history $\mathcal{H}$, and accesses the sampling oracle $EX_{\mathcal{C}}()$ of a given covertext channel $\mathcal{C}$ and returns a stegotext $S \in \Sigma^\ell$;*
- *$SD = SD(K, S, \mathcal{H})$ is the decoding algorithm that takes K, S, and $\mathcal{H}$, and having access to the sampling oracle $EX_{\mathcal{C}}()$ returns a message M'.*

$\mathcal{S}$ is called a black-box stegosystem *if SE and SD have no a priori knowledge about the distribution of the covertext channel (except its min-entropy) and can obtain information about it only by querying the sampling oracle*[1].

The key K is shared by Alice and Bob beforehand and is kept secret from an adversary. All further actions of Alice are specified by *SE*, those of Bob by *SD*. The time complexities of the algorithms *SE*, *SD* are measured with respect to n, κ, and σ, where an oracle query is charged as one unit step. A stegosystem is *computationally efficient* if its time complexities are polynomially bounded. By convention, the running time of an algorithm includes the so-called *description size* of that algorithm with respect to some standard encoding.

Ideally, one would expect that the decoder always succeeds in extracting the original message M from the stegotext. Since this may not always be possible, we define for $M \in \{0,1\}^n$ the unreliability of a stegosystem.

Definition 3 (Unreliability). *Given a covertext channel $\mathcal{C}$, the unreliability of $\mathcal{S}$ is given by*

$$\mathtt{UnRel}_{\mathcal{C},\mathcal{S}} \;:=\; \max_{\mathcal{H}} \max_{M \in \{0,1\}^n} \Pr_{K \in_R \{0,1\}^\kappa}[SD(K, SE(K, M, \mathcal{H}), \mathcal{H}) \neq M] \;.$$

Before analysing different notions of security in steganography in the computational setting, we have to define the scenario of steganographic analysis. An adversary, called the warden W, tries to find out whether or not the communicating parties are using steganography. This is the standard model of a (passive) adversary. Let $SE(K, M, \mathcal{H})$ with access to $EX_{\mathcal{C}}(\mathcal{H})$ be denoted by $SE^{\mathcal{C}}(K, M, \mathcal{H})$. We define the oracle OC that for a given message $M \in \{0,1\}^n$ and channel history $\mathcal{H}$ returns a truly random covertext $c_1 c_2 \ldots c_\ell$ of length $\ell = |SE^{\mathcal{C}}(K, M, \mathcal{H})|$ from $\mathcal{C}$ with history $\mathcal{H}$, i.e., each c_i is drawn according to $D_{\mathcal{H}||c_1 c_2 \ldots c_{i-1}}$.

Definition 4 (Advantage of a warden performing a chosen hiddentext attack). *A probabilistic algorithm W is a* **(t, q, λ)-warden** *for the stegosystem $\mathcal{S} = [SE, SD]$ if*

- *W runs in time t and accesses a reference oracle $EX_{\mathcal{C}}()$ that he can query for samples from the covertext channel $\mathcal{C}$ with a history $\mathcal{H}$;*
- *W can make q queries of total bit length λ to a challenge oracle CH which is either $SE^{\mathcal{C}}(K, M, \mathcal{H})$ or $OC(M, \mathcal{H})$, where M of length n and $\mathcal{H}$ can be chosen by W;*
- *the task of W is to determine the use of the stegosystem $\mathcal{S}$ with the help of the challenge oracle: $W^{\mathcal{C},CH} = 1$ means that W decides on "stegotext", resp. $W^{\mathcal{C},CH} = 0$ on "covertext".*

The advantage *of W for a stegosystem $\mathcal{S}$ using a covertext channel $\mathcal{C}$ is defined as*

$$\mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \;:=\; \Pr_{K \in_R \{0,1\}^\kappa}[W^{\mathcal{C}, SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr[W^{\mathcal{C}, OC(\cdot,\cdot)} = 1] \;. \tag{1}$$

[1] As usual, we assume that for each legal history $\mathcal{H}$ the encoding or decoding algorithm can query an arbitrary number of samples from the covertext channel with history $\mathcal{H}$.

Note that for technical reasons we do not take the absolute value of the difference of probabilities, thus a bad warden may even have a negative advantage. By complementing the decision of such a bad W we get another warden which achieves the same positive amount of advantage. Since the security measures considered here are always based on the best warden negative advantages have no influence.

For maximising the advantage, W may depend on the channel $\mathcal{C}$. In the most favourable case, W may possess a complete specification of $\mathcal{C}$, so that he even does not need to query the reference oracle. Such information about $\mathcal{C}$ is part of the description of W.[2] This makes the adversary extremely powerful in the black-box stegosystem setting.

Random Permutations

Below we recall some notions from cryptography required for the specification of the encoding function *SE*. Let $F : \{0,1\}^k \times \{0,1\}^l \to \{0,1\}^L$ be a function. Here $\{0,1\}^k$ is considered as the key space of F. For each key $K \in \{0,1\}^k$ we define the subfunction $F_K : \{0,1\}^l \to \{0,1\}^L$ by $F_K(x) = F(K,x)$. F is called a family of permutations if $l = L$ and for each key K the subfunction F_K is a permutation on $\{0,1\}^l$. Let $PERM(l)$ denotes the family of all permutations on $\{0,1\}^l$.

Following [3] we define a security notion for pseudorandom permutations with the help of a distinguisher, who is comparable to a warden for detecting steganography.

Definition 5. *For a family F of permutations the* advantage of a probabilistic distinguisher D *having access to a challenge oracle that returns either values according to F_K for unknown K or according to a random permutation P is given by*

$$\texttt{PRP-Adv}_F(D) \;=\; \Pr_{K \in_R \{0,1\}^k}[D^{F_K(\cdot)} = 1] - \Pr_{P \in_R PERM(l)}[D^{P(\cdot)} = 1],$$

The **insecurity of a family of permutations F** *is defined as*

$$\texttt{PRP-InSec}_F(t,q) \;:=\; \max_D \texttt{PRP-Adv}_F(D)\ ,$$

where the maximum is taken over all probabilistic distinguishers D running in at most t steps and making at most q oracle queries.

A sequence $\{F_k\}_{k \in \mathbb{N}}$ is called pseudorandom *if for all polynomially bounded D, $\texttt{PRP-Adv}_F(D)$ is negligible in k.*

For our constructions given below we assume the existence of families of pseudorandom functions.

3 Security Levels of Stegosystems

In this section different notions of security will be discussed and their specific strength for both opponents in the game will be investigated. We consider arbitrary restricted families $\mathcal{F}$ of covertext channels instead of simply *all* channels over the alphabet Σ, which has typically been done in theoretical studies so far (few papers have studied specific families, like e.g. memoryless channels [13]).

[2] In contrast to [5] we do not explicitly mention the *description size of the warden*, but assume this to be included in the running time t (W has to read this information at least once).

Restricting the set of channels allows a finer differentiation and models the practical situation in steganography and steganalysis better. For example, any specifically designed stegosystem $\mathcal{S}$ for embedding hidden information in digital images is likely to give the stegoanalyst a much better advantage when used for other channels that deviate significantly from images like, e.g. music. But this property seems to be useless for a stegoanalyst if $\mathcal{S}$ is never used other than for images. Commonly used as an (in)security measure, see e.g. [9, 5], is the following quantity.

Definition 6. *The* **insecurity of a stegosystem $\mathcal{S}$** *with respect to a channel family $\mathcal{F}$ is defined as follows, where for given complexity bounds (t, q, λ) we take into account all (t, q, λ)-wardens W:*

$$\mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \;:=\; \max_{W} \max_{\mathcal{C}\in\mathcal{F}} \; \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \;\; .$$

The security of a system $\mathcal{S}$ with respect to $\mathcal{F}$ is defined as $1 - \mathtt{InSec}_{\mathcal{F},\mathcal{S}}$. If a system $\mathcal{S}$ generates a small value for $\mathtt{InSec}_{\mathcal{F},\mathcal{S}}$ then it achieves the highest security level: For every channel from the family no warden can detect the stegosystem with a significant advantage. Thus, $\mathcal{S}$ is a good *universal* system for $\mathcal{F}$. However, currently no secure and efficient stegosystems are known for any non-trivial channel family. Even more, it has been proven that for a specific simple family of channels such universal systems do not exist [5]. But does this result mean that the warden can control steganography for such channel families? The problem is that if a stegosystem $\mathcal{S}$ is *insecure*, then there *exists* a single channel $\mathcal{C}_0$ in $\mathcal{F}$ such that some specific strategy W_0 can detect steganography over $\mathcal{C}_0$. However, this does *not* imply that the warden can detect the usage of the stegosystem $\mathcal{S}$ for any other channel in $\mathcal{F}$. Therefore the above measure does not fit well from the point of view of a steganalyst: an insecure stegosystem $\mathcal{S}$ can remain undetectable for almost all channels in $\mathcal{F}$. One could modify the above definition in a natural way such that it reflects the necessities of steganalysis.

Definition 7. *The* **channel-universal detectability** *of a stegosystem $\mathcal{S}$ with respect to a channel family $\mathcal{F}$ is defined as follows, where the maximum is taken over all (t, q, λ)-wardens W:*

$$\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \;:=\; \max_{W} \min_{\mathcal{C}\in\mathcal{F}} \; \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \;\; .$$

If a stegosystem $\mathcal{S}$ is channel-universally detectable with respect to the family $\mathcal{F}$, then using some universal strategy W can detect the usage of the stegosystem $\mathcal{S}$ for many other channels $\mathcal{C}$ in $\mathcal{F}$. This guarantees the highest detectability level. But it is unclear how such a level of detectability can be achieved. Moreover, if for some stegosystem $\mathcal{S}$ the value $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}$ is small, one cannot guarantee that $\mathcal{S}$ is secure for *every* channel in $\mathcal{F}$. One may construct a stegosystem $\mathcal{S}_0$ that works well for only one channel $\mathcal{C}_0 \in \mathcal{F}$ – yielding a small value $\mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C}_0,\mathcal{S}_0}(W)$. Such a stegosystem is not channel-universally detectable since for $\mathcal{C}_0$ *no* strategy of the warden is able to detect $\mathcal{S}_0$ with a significant advantage. But the system can still be easily detectable for most other channels in $\mathcal{F}$.

Thus, for a security analysis it is extremely important who selects the covertext channel – the encoder or the warden. For most applications it seems unrealistic to assume that the warden can dictate to the encoder which covertext channel to use. In case that neither opponent has a free choice, one should take into account how much knowledge about the covertext distribution each one is given a priori (see Fig. 1). This may be helpful despite the sampling oracle.

Let us summarize the discussion so far as follows. For any channel family $\mathcal{F}$ and for every stegosystem $\mathcal{S}$ and all t, q, λ it holds:

$$0 \;\le\; \mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \;\le\; \mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \;\le\; 1 \;\; .$$

For most non-trivial families $\mathcal{F}$ and reasonable stegosystems $\mathcal{S}$ one typically observes that $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}$ is small and $\mathtt{InSec}_{\mathcal{F},\mathcal{S}}$ is large. But in such a case we are not able to provide any reasonable degree of insecurity/detectability of the system. Our goal will be to give and to analyse more appropriate measures for insecurity/detectability of stegosystems. From the definition of *channel-universal* detectability it is natural to derive *channel-specific* detectability, which we define as follows.

Definition 8. *The* **channel-specific detectability** *of a stegosystem $\mathcal{S}$ with respect to a channel family $\mathcal{F}$ is defined as follows with the maximum taken over all (t, q, λ)-wardens W:*

$$\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \ := \ \min_{\mathcal{C}\in\mathcal{F}} \max_{W} \ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \ .$$

From the order of the min- and max-operators we get immediately

$$\max_{W} \min_{\mathcal{C}\in\mathcal{F}} \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \ \leq \ \min_{\mathcal{C}\in\mathcal{F}} \max_{W} \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \ \leq \ \max_{\mathcal{C}\in\mathcal{F}} \max_{W} \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \ = \ \max_{W} \max_{\mathcal{C}\in\mathcal{F}} \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \ ,$$

which implies

Lemma 1. *For every channel family $\mathcal{F}$ and stegosystem $\mathcal{S}$ and parameters t, q, λ it holds:*

$$\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \ \leq \ \mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \ \leq \ \mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) \ .$$

Now, if the value of $\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}$ is large, then for every channel $\mathcal{C}$ in $\mathcal{F}$ there exists some warden which can detect the use of steganography for this particular channel $\mathcal{C}$ by exploiting his specific strategy W. This definition relaxes the strong assumption of universality with respect to the covertext channel in use. However, while each W might work well for his particular $\mathcal{C}$, W may perform poorly on all other channels of $\mathcal{F}$. Thus, in contrast to a high value for $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}$, giving the warden good confidence in his power, a high value of $\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}$ does not really say much about the power of a warden, unless he knows Alice's choice of a channel. On the other hand, for a small value of $\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}$ the stegosystem $\mathcal{S}$ may work well for most channels in $\mathcal{F}$.

It should be apparent that a different security definition is desirable which takes into account that neither the warden nor the steganographer may be universal for *all* channels in $\mathcal{F}$, but perhaps still be able to perform well on average. Therefore, assuming a probability distribution of channels $\mathcal{C}$ in the family $\mathcal{F}$, we will generalise the notion of advantage given in (1) from a fixed channel to a channel chosen at random:

$$\begin{aligned}\mathtt{Adv}^{\mathsf{cha}}_{\mathcal{F},\mathcal{S}}(W) &:= \Pr_{\mathcal{C}\in_R\mathcal{F}} \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \\ &= \Pr_{\mathcal{C}\in_R\mathcal{F}, K\in_R\{0,1\}^\kappa}[W^{\mathcal{C},SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C}\in_R\mathcal{F}}[W^{\mathcal{C},OC(\cdot,\cdot)} = 1] \ .\end{aligned}$$

Definition 9. *The* **detectability on average** *of a stegosystem $\mathcal{S}$ with respect to the channel family $\mathcal{F}$ is given as follows, where the maximum is taken over all (t, q, λ)-wardens W:*

$$\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}(t, q, \lambda) : := \ \max_{W} \ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{F},\mathcal{S}}(W) \ .$$

This definition has clear advantages over the previous ones. If for a stegosystem $\mathcal{S}$ the value of $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}$ is low, then Alice can be assured that W in most cases will not be able to detect steganography, whereas a high value indicates that W is likely to catch her. Thus, $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}$

provides a measure that can be used by both Alice and W to assess their expected performance in the game if neither has complete control over the channel. It should be noted that a family of channels $\mathcal{F}$ with a distribution on the family cannot simply be aggregated to a single, more complicated channel $\mathcal{C}_{\mathcal{F}}$. This would be a quite different situation for Alice and Warden.[3]

This new measure for insecurity/detectability of stegosystems corresponds better to real life intuition of insecurity than the commonly used definition. In fact, in real life steganalysis, our approach is already implicitly used in empirical analyses of particular stegosystems. For example, it is not difficult to see that the steganographic algorithm F5 used to embed hidden information in JPEG images [15] is insecure with respect to the common insecurity definition. But this observation seems to be useless to a stegoanalyst, for whom a much more appropriate approach to analyse the insecurity would be to use the new definition and consider a universal algorithm to detect the use of F5, like it was done e.g. in [6]. In this example one could specify formally $\mathcal{F}$ as a family of channels $\mathcal{C}_\omega$ of JPEG-compressed images of different scenes or taken by different types of digital cameras specified by ω.

In the rest of this paper, we will discuss and analyse scenarios showing that $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}$ is indeed much better suited than the other security notions. Detectability on average is related to the previously defined security measures as follows (cf. Fig. 1):

Lemma 2. *For every channel family $\mathcal{F}$, every stegosystem $\mathcal{S}$ and all t, q, λ it holds:*

$$\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}(t,q,\lambda) \;\leq\; \mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}(t,q,\lambda) \;\leq\; \mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t,q,\lambda)\ .$$

Proof. If $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}}(t,q,\lambda) = \min_{\mathcal{C}\in\mathcal{F}} \max_W \ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) = \epsilon$ there must be a warden W_0 with bounds t, q, λ that achieves an advantage of at least ϵ for every channel in $\mathcal{F}$. For this warden and any probability distribution μ on $\mathcal{F}$ its expected advantage $\sum_C \mu(C)\ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W_0)$ will be at least ϵ. Thus,

$$\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}(t,q,\lambda) \;=\; \max_W \sum_C \mu(C)\ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \;\geq\; \sum_C \mu(C)\ \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W_0) \;\geq\; \epsilon\ .$$

Furthermore, the average advantage over $\mathcal{F}$ is upper-bounded by the maximum advantage, thus $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}(t,q,\lambda) \;\leq\; \mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t,q,\lambda)$.

□

From our analysis given below it follows that $\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}}$ and $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}}$ are incomparable. By construction a specific family of channels we will show:

Theorem 1. *There exists a channel family $\mathcal{F}$ and stegosystems $\mathcal{S}_{\mathcal{F}}$ and $\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}$ such that for appropriate parameters t, q, λ:*

$$\begin{aligned} \mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}(t,q,\lambda) &\ll \mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}(t,q,\lambda) \quad \textit{and} \\ \mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}}(t,q,\lambda) &\ll \mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}}(t,q,\lambda)\ . \end{aligned}$$

This property will follow from the bounds shown in Theorems 4 to 7.

[3] For example, a family of quite restricted channels may aggregate to the uniform distribution on the document space. For such a channel Alice can achieve perfect secure steganography easily using a random permutation, whereas this is not guaranteed for the individual channels.

4 Undetectable Stegosystems

In steganography there are two extreme cases of channel families: in the first case, using the sampling oracle the encoder can obtain full knowledge about the covertext distribution; in the second case it is extremely difficult for the encoder to deduce anything about the covertext distribution. For families of the first type secure stegosystems (with low $\mathtt{InSec}$) can be built. This is not possible for the second type of families, since the encoder cannot even perform some simple tests for the constructed stegotext, whereas, according to the definition of $\mathtt{InSec}$, the warden can have full knowledge about the covertext distribution used. In this section we show that the situation changes drastically if a symmetry in knowledge about the channel is given to both opponents. In particular, we prove that it is possible to construct undetectable stegosystems if it is difficult to deduce something about the covertext distribution. We construct a stegosystem $\mathcal{S}_{\mathcal{F}}$ that works for a given channel family $\mathcal{F}$, i.e., we assume Alice and Bob know that a fixed covertext channel $\mathcal{C}$ is chosen from $\mathcal{F}$, but they have no additional knowledge about $\mathcal{C}$. Thus, although the system is not universal for *all* channels, it is universal for all channels in the family $\mathcal{F}$. The system works for families $\mathcal{F}$ of channels with finite descriptions and efficiently computable distribution functions defined as follows.

Definition 10. *Let $\mathcal{F}$ be a family of channels $\mathcal{C}_\omega$ indexed by strings $\omega \in \{0,1\}^\star$. These channels share a document space Σ that has an arbitrary linear ordering "$\leq$", e.g. lexicographic order. $D^\omega_{\mathcal{H}}$ denotes the probability distribution of the channel $\mathcal{C}_\omega$ with respect to history $\mathcal{H}$, i.e. $\Pr_{D^\omega_{\mathcal{H}}}[x]$ is the probability that document x is generated by $\mathcal{C}_\omega$ with history $\mathcal{H}$. The (cumulative) distribution functions of $\mathcal{F}$ defined by $F^\omega_{\mathcal{H}}(c) := \sum_{x \leq c} \Pr_{D^\omega_{\mathcal{H}}}[x]$ are called efficiently computable if there exists a polynomially time bounded algorithm that on input ω, $\mathcal{H}$ and c outputs $F^\omega_{\mathcal{H}}(c)$.*

4.1 Interval Encoding

Assume that we want to encode b bits. We number the bitstrings from 0 to $2^b - 1$ and consider the ρ-th bitstring. To encode ρ we can use all documents c with a value $F^\omega_{\mathcal{H}}(c)$ in the interval $I_\rho :=]\rho \cdot 2^{-b}, (\rho+1) \cdot 2^{-b}]$. Next we choose a random number z_ρ in this interval and select among all documents with positive probability $\Pr_{D^\omega_{\mathcal{H}}}[c]$ the minimum c such that $z_\rho \leq F^\omega_{\mathcal{H}}(c)$. Let us denote this mapping by $\mathbf{IntervalEncode}(\omega, \mathcal{H}, \rho)$. If we first select a value ρ uniformly at random and then apply IntervalEncode$(\omega, \mathcal{H}, \rho)$, it is guaranteed that each document $c \in \Sigma$ is chosen with probability exactly $\Pr_{D^\omega_{\mathcal{H}}}[c]$, thus we generate the same distribution as $\mathcal{C}_\omega$.

The decoding works as follows. Receiving document c, Bob computes the value ρ' such that $F^\omega_{\mathcal{H}}(c) \in I_{\rho'}$. The value ρ' differs from the correct value ρ if in the encoding of $z_\rho \in I_\rho$ there was no document c' with $z_\rho \leq F^\omega_{\mathcal{H}}(c')$ such that $F^\omega_{\mathcal{H}}(c')$ belongs to I_ρ, too. In other words, z_ρ is a value such that in the interval $[z_\rho, (\rho+1) \cdot 2^{-b}]$ the distribution function $F^\omega_{\mathcal{H}}$ does not increase. If the min-entropy of the channels in $\mathcal{F}$ is at least h, in any subinterval of I_ρ of length at least 2^{-h} $F^\omega_{\mathcal{H}}$ must be strictly increasing. The probability to select a *bad* z_ρ is therefore less than $2^{-(h-b)}$. Thus, for each of the b bits, the probability that this bit is incorrectly decoded is bounded by $2^{-(h-b)}$.

4.2 A Strong Private Key Stegosystem

The stegosystem $\mathcal{S}_{\mathcal{F}}$ specified in Fig. 2 uses this interval coding technique . Recall that $\ell = n/b$ is an integer specifying the number of blocks into which a message M is split. To encrypt a message M, we use families of pseudorandom permutations $\mathtt{PRP} : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ to spread M

uniformly. The private key $K = \omega||K_1 \in \{0,1\}^{\eta+k}$ for encoder and decoder is chosen uniformly at random. The prefix ω of the key K is a random string of length η that is used to select a random element $\mathcal{C}_\omega$ of $\mathcal{F}$. The length η depends on the Family $\mathcal{F}$. The suffix K_1 is used to specify which pseudorandom permutation $\mathtt{PRP}(K_1, \cdot) = \mathtt{PRP}_{K_1}(\cdot)$ is selected.

Procedure $\mathcal{S}_\mathcal{F}$**-Encode**(K, $\mathcal{H}$, M)
Input: private key $K = \omega||K_1$; history $\mathcal{H}$;
hiddentext $M \in \{0,1\}^n$
choose $T_0 \in_R \{0,1\}^n$;
let $T_1 := \mathtt{PRP}_{K_1}(T_0 \oplus M)$;
parse T_0T_1 into $u_1u_2 \dots u_{2\ell}$, where $|u_i| = b$;
for $j := 1, \dots, 2\ell$ **do**
  let ρ_j be the integer with binary repr. u_j;
  $s_j :=$ **IntervalEncode**$(\omega, \mathcal{H}, \rho_j)$;
  $\mathcal{H} := \mathcal{H}||s_j$;
Output: $s_1s_2 \dots s_{2\ell}$

Procedure $\mathcal{S}_\mathcal{F}$**-Decode**(K, $\mathcal{H}$, s)
Input: private key $K = \omega||K_1$; history $\mathcal{H}$;
stegotext $s = s_1, \dots, s_{2\ell}$;
for $j := 1, \dots, 2\ell$ **do**
  determine ρ_j s.t. $\frac{\rho_j}{2^b} < F^\omega_\mathcal{H}(s_j) \leq \frac{\rho_j+1}{2^b}$;
  let u_j be the b-bit binary repr. of ρ_j;
let $T_0 := u_1 \dots u_\ell$ and $T_1 := u_{\ell+1} \dots u_{2\ell}$;
$M := \mathtt{PRP}^{-1}_{K_1}(T_1) \oplus T_0$;
Output: M

Fig. 2. The stegosystem $\mathcal{S}_\mathcal{F}$ based on interval encoding of a random channel

The crucial property of $\mathcal{S}_\mathcal{F}$ is that the random choice of ω for the channel $\mathcal{C}_\omega$ is independent of the *real* channel $\mathcal{C}$ generating the covertexts. Alice and Bob just randomly select a channel to work with, knowing that with high probability it is a wrong one. For this reason, the stegosystem $\mathcal{S}_\mathcal{F}$ may output samples that are not in the support of $\mathcal{C}$, which may make it insecure for many families $\mathcal{F}$. However, we will show that this system is not channel-universally detectable since the correct channel may be picked by chance.

4.3 Distinguishing Channels

Below we will describe a new framework for analysing the security of stegosystems in a realistic environment where covertext channels are not completely known to any opponent. Security is based on the hardness for distinguishing channels of a channel family $\mathcal{F}$.

Definition 11 (Channel distinguisher).
A probabilistic algorithm Q is a (t, q, λ)**-distinguisher** *for a channel family $\mathcal{F}$ if*

- *Q runs in time t and accesses a* **reference oracle** *$EX_\mathcal{C}()$, for some covertext channel $\mathcal{C} \in \mathcal{F}$, which it can query for samples from $\mathcal{C}$ with a history $\mathcal{H}$ that can be chosen by Q;*
- *Q can make a number of q queries of total bit length λ to a* **challenge oracle** *CH which is either $EX_\mathcal{C}()$ or $EX_{\mathcal{C}'}()$ for some other covertext channel $\mathcal{C}' \in \mathcal{F}$;*
- *Q has to determine whether the channel defining the challenge oracle CH is the same as the channel $\mathcal{C}$ of the reference oracle $EX_\mathcal{C}()$ formalized as: $Q^{\mathcal{C},CH}$ outputs 1 if he thinks that they differ, whereas $Q^{\mathcal{C},CH} = 0$ means that they are identical.*

The **distinguishability for a channel family** $\mathcal{F}$ *is defined as follows, where the maximum is taken over all (t, q, λ)-distinguishers Q and $\mathcal{C}, \mathcal{C}' \in_R \mathcal{F}$ are chosen independently:*

$$\mathtt{Dist}_\mathcal{F}(t, q, \lambda) \;:=\; \max_Q \Pr_{\mathcal{C},\mathcal{C}' \in_R \mathcal{F}}[Q^{\mathcal{C},\mathcal{C}'} = 1] - \Pr_{\mathcal{C} \in_R \mathcal{F}}[Q^{\mathcal{C},\mathcal{C}} = 1] \;.$$

If it is infeasible to distinguish two random elements from $\mathcal{F}$ then Alice, of course, has a problem to find out the real channel. She may either guess a document in Σ and hope that it is in the support of the real channel, or she may ask for a number of covertexts that is (on average) exponential in b, until she gets one that codes the hiddentext M. But the adversary faces the same problem to determine the correct channel unless this information is directly given to him, which seems unrealistic in practice. The following theorem establishes a tight relationship between the distinguishability of a channel family $\mathcal{F}$ and detectability on average for the above stegosystem applied to $\mathcal{F}$. To shorten the notion define

$$\xi(\lambda, n) \ := \ \left(\frac{\lambda^2}{n^2} - \frac{\lambda}{n}\right) \cdot 2^{-n} \ .$$

Theorem 2. *Let $\mathcal{F}$ be a family of channels $\mathcal{C}_\omega$ (indexed by strings $\omega \in \{0,1\}^\eta$) over an alphabet Σ of size 2^σ with efficiently computable distribution functions and min-entropy h larger than the rate b. The elements of $\mathcal{F}$ are selected uniformly at random as covertext channels. Then the stegosystem $\mathcal{S}_\mathcal{F}$ described above achieves rate b, unreliability bounded by $n2^{-(h-b)}$, and runs in time polynomial with respect to η, σ, and the message length n. Furthermore, there is a fixed polynomial p such that*

$$\begin{aligned}
\texttt{AvgDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}}(t,q,\lambda) \ &\geq \ \texttt{Dist}_\mathcal{F}(t,q,\lambda) - 2\,\texttt{PRP-InSec}_{\texttt{PRP}}(p(t),\lambda/n) - \xi(\lambda,n) \ , \\
\texttt{AvgDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}}(t,q,\lambda) \ &\leq \ \texttt{Dist}_\mathcal{F}(p(t),q,\lambda) + 2\,\texttt{PRP-InSec}_{\texttt{PRP}}(p(t),\lambda/n) + \xi(\lambda,n) \ .
\end{aligned}$$

These bounds show that the average detectability of the stegosystem $\mathcal{S}_\mathcal{F}$ is basically identical to the distinguishability of the channel family modulo the insecurity of the pseudorandom permutations.

Proof. To show the upper bound let W be a (t,q,λ)-warden of maximum average advantage, that means

$$\texttt{AvgDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}}(t,q,\lambda) \ = \ \Pr_{\mathcal{C},K}[W^{\mathcal{C},SE^\mathcal{C}(K,\cdot,\cdot)} = 1] - \Pr_\mathcal{C}[W^{\mathcal{C},\mathcal{C}} = 1] \ .$$

Recall that W uses a reference oracle for a random channel $\mathcal{C}$ from $\mathcal{F}$ and as challenge oracle either an oracle for $SE^\mathcal{C}$, the encoding procedure $\mathcal{S}_\mathcal{F}$-**Encode** working with covertext channel $\mathcal{C}$, or simply an oracle for $\mathcal{C}$ itself. We bound the advantage of W for differentiating between $SE^\mathcal{C}$ and $\mathcal{C}$ indirectly by considering a random channel in between. Let $W^{\mathcal{C},\mathcal{C}_\omega}$ denote W with oracles for the channels $\mathcal{C}$ and $\mathcal{C}_\omega$. The oracle for $\mathcal{C}_\omega$, given message $M \in \{0,1\}^n$ and history $\mathcal{H}$, returns a truly random sequence $c_1 c_2 \ldots c_{2\ell}$ of length $2\ell = |SE^\mathcal{C}(K,M,\mathcal{H})|$ from $\mathcal{C}_\omega$ with history $\mathcal{H}$. This way we can expand the formula to

$$\begin{aligned}
&\texttt{AvgDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}}(t,q,\lambda) \ = \\
&\qquad \Pr_{\mathcal{C},K}[W^{\mathcal{C},SE^\mathcal{C}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1] \ + \ \Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1] - \Pr_\mathcal{C}[W^{\mathcal{C},\mathcal{C}} = 1] \ .
\end{aligned}$$

For a suitable polynomial p, a $(p(t),q,\lambda)$-distinguisher $Q^{\mathcal{C},CH}$ for the channel family $\mathcal{F}$ can simply be obtained by simulating the warden W with challenge oracle either $\mathcal{C}$ or $\mathcal{C}_\omega$. Thus,

$$\Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1] - \Pr_\mathcal{C}[W^{\mathcal{C},\mathcal{C}} = 1] \ \leq \ \texttt{Dist}_\mathcal{F}(p(t),q,\lambda) \ ,$$

since the probability distribution of ω over descriptions of channels in $\mathcal{F}$ is equal to the probability distribution $\mathcal{C} \in_R \mathcal{F}$. It remains to bound

$$\Pr_{\mathcal{C},K}[W^{\mathcal{C},SE^\mathcal{C}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1] \ .$$

Let PRP be the family of pseudorandom permutations used in the stegosystem $\mathcal{S}_{\mathcal{F}}$, and $CBC[\mathtt{PRP}]$ the symmetric encryption scheme with encryption procedure $\mathcal{E}_{K_1}$ and decryption procedure $\mathcal{D}_{K_1}$, defined in Fig. 3, where K_1 is a secret key.

Procedure $\mathcal{E}_{K_1}(M)$	**Procedure** $\mathcal{D}_{K_1}(T)$
Input: plaintext $M \in \{0,1\}^n$	**Input**: ciphertext $T \in \{0,1\}^{2n}$
$T_0 \in_R \{0,1\}^n$;	parse T as $T_0 \| T_1$;
$T_1 := \mathtt{PRP}_{K_1}(T_0 \oplus M)$;	$M := \mathtt{PRP}^{-1}_{K_1}(T_1) \oplus T_0$;
Output: $T = T_0 \| T_1$	**Output**: M

Fig. 3. A symmetric cryptosystem

We will apply $(\mathcal{E}_{K_1}, \mathcal{D}_{K_1})$ to simulate the encryption and decryption of messages M used by the stegosystem $\mathcal{S}_{\mathcal{F}}$.

The *real-or-random* insecurity $\mathtt{ES\text{-}InSec}^{\mathsf{ror}}_{\mathcal{ES}}(t', q', \lambda')$ of an encryption scheme $\mathcal{ES} = (\mathcal{E}_{K_1}, \mathcal{D}_{K_1})$ is defined as the maximum advantage $\mathtt{ES\text{-}Adv}^{\mathsf{ror}}_{\mathcal{ES}}(A)$ over all probabilistic adversaries A running in at most t' steps and making at most q' oracle queries of total length λ', where the advantage is given by

$$\mathtt{ES\text{-}Adv}^{\mathsf{ror}}_{\mathcal{ES}}(A) \;=\; \Pr_{K_1}[A^{\mathcal{E}_{K_1}(\cdot)} = 1] - \Pr_{K_1}[A^{\mathcal{E}_{K_1}(\$)} = 1] \;.$$

Here, the (real encryption) oracle $\mathcal{E}_{K_1}(\cdot)$ on input M, returns $\mathcal{E}_{K_1}(M)$, while the (random) oracle $\mathcal{E}_{K_1}(\$)$ on input M, returns $\mathcal{E}_{K_1}(r)$ with $r \in_R \{0,1\}^{|M|}$.

In [3] the following bound on the *real-or-random* insecurity of a system like $CBC[\mathtt{PRP}]$ has been shown:

$$\mathtt{ES\text{-}InSec}^{\mathsf{ror}}_{CBC[\mathtt{PRP}]}(t', q', \lambda') \;\leq\; 2 \cdot \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(t'', q'') + \xi(\lambda', n) \;, \tag{2}$$

with $t'' = t' + c\lambda'$ for some constant c and $q'' = \lambda'/n$. $\xi(x, n) \;:=\; \left(\frac{x^2}{n^2} - \frac{x}{n}\right) \cdot 2^{-n}$

Using the warden W, we will now design an adversary A against the symmetric encryption scheme $CBC[\mathtt{PRP}]$ working as follows. First, A chooses a random covertext channel $\mathcal{C}$ and a random private key $K = \omega \| K_1$. Then it simulates the computation of W with reference oracle $\mathcal{C}$ and challenge oracle either $SE^{\mathcal{C}}(K, \cdot, \cdot)$ or $OC(\cdot, \cdot)$.
Whenever W tries to query the challenge oracle CH with M and $\mathcal{H}$, A does the following:

1. it queries its oracle for $\mathcal{E}_{K_1}$ with M;
2. with the answer $\hat{T}_0\hat{T}_1$, A simulates the procedure $\mathcal{S}_{\mathcal{F}}$-**Encode** with key ω, history $\mathcal{H}$, but skips the computation of T_0T_1 and sets the string T_0T_1 to $\hat{T}_0\hat{T}_1$;

A passes the output $s_1 \ldots s_{2\ell}$ of the simulation as an answer of the challenge oracle to W.
Finally, A returns the output value of W.

If W obeys the complexity bounds (t, q, λ) then A can work in time $p'(t)$ for some polynomial p' depending on the complexity of the pseudorandom permutations and the evaluation of the distribution functions. It needs at most q oracle questions. Since the stegosystem $\mathcal{S}_{\mathcal{F}}$ uses the encryption scheme $(\mathcal{E}_{K_1}, \mathcal{D}_{K_1})$ the probabilities $\Pr_{K_1}[A^{\mathcal{E}_{K_1}(\cdot)} = 1]$ and $\Pr_{\mathcal{C}, K=\omega\|K_1}[W^{\mathcal{C}, SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1]$ are equal.

Similarly, $\Pr_{K_1}[A^{\mathcal{E}_{K_1}(\$)} = 1] = \Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1]$. Then, using the estimation (2), for an appropriate polynomial p we can conclude

$$\begin{aligned}
&\Pr_{\mathcal{C},K=\omega||K_1}[W^{\mathcal{C},SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C},\omega}[W^{\mathcal{C},\mathcal{C}_\omega} = 1] \\
&\quad = \Pr_{K_1}[A^{\mathcal{E}_{K_1}(\cdot)} = 1] - \Pr_{K_1}[A^{\mathcal{E}_{K_1}(\$)} = 1] \\
&\quad = \mathtt{ES\text{-}Adv}^{\mathsf{ror}}_{CBC[\mathtt{PRP}]}(A) \;\leq\; \mathtt{ES\text{-}InSec}^{\mathsf{ror}}_{CBC[\mathtt{PRP}]}(p'(t), q, \lambda) \\
&\quad \leq\; 2 \cdot \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t), \lambda/n) + \xi(\lambda, n) \;.
\end{aligned}$$

This completes the proof of the upper bound.

To prove the lower bound

$$\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}(t, q, \lambda) \;\geq\; \mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) - 2 \cdot \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t), \lambda/n) - \xi(\lambda, n)$$

first note that the last estimation does not only hold for W, but for any adversary Q with the same complexity bounds (t, q, λ), thus

$$\Pr_{\mathcal{C},K=\omega||K_1}[Q^{\mathcal{C},SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C},\omega}[Q^{\mathcal{C},\mathcal{C}_\omega} = 1] \;\leq\; 2 \cdot \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t), \lambda/n) + \xi(\lambda, n) \;.$$

Let Q be a (t, q, λ)-distinguisher such that

$$\mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) \;=\; \Pr_{\mathcal{C},\omega}[Q^{\mathcal{C},\mathcal{C}_\omega} = 1] - \Pr_{\mathcal{C}}[Q^{\mathcal{C},\mathcal{C}} = 1] \;.$$

We can split this advantage – now with the help of an oracle for $SE^{\mathcal{C}}$ – into

$$\begin{aligned}
&\mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) \;\leq \\
&\qquad \Pr_{\mathcal{C},\omega}[Q^{\mathcal{C},\mathcal{C}_\omega} = 1] - \Pr_{\mathcal{C},K}[Q^{\mathcal{C},SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] \;+\; \Pr_{\mathcal{C},K}[Q^{\mathcal{C},SE^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C}}[Q^{\mathcal{C},\mathcal{C}} = 1] \;.
\end{aligned}$$

In the second term Q acts like a (t, q, λ)-bounded warden, thus his advantage is bounded by $\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}(t, q, \lambda)$. This gives

$$\mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) \;\leq\; 2 \cdot \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t), \lambda/n) + \xi(\lambda, n) + \mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}(t, q, \lambda) \;.$$

□

5 Insecurity versus Detectability

In [5] for a specific family $\mathcal{F} = \mathtt{PRC}_\eta$ of covertext channels, called pseudorandom flat h-channels, the following result is shown, where the parameter η describes the length of a random seed and h the entropy of the channels.

> For every stegosystem $\mathcal{S}$ of small unreliability $\mathtt{UnRel}_{\mathcal{F},\mathcal{S}}$ and small insecurity $\mathtt{InSec}_{\mathcal{F},\mathcal{S}}(t, q, \lambda)$, for polynomially bounded t, q, λ, there exists a channel $\mathcal{C}$ in $\mathcal{F}$ such that the query complexity of $\mathcal{S}$ has to be large.

This implies that a secure, reliable and efficient stegosystem does not exist for this channel family – for every efficient stegosystem $\mathcal{S}$ the value $\mathtt{InSec}_{\mathcal{F},\mathcal{S}}$ is large if Alice has to fight against arbitrary polynomially bounded wardens. Obviously, one can conclude that for every channel family $\mathcal{F}'$ that includes pseudorandom flat h-channels $\mathtt{PRC}_\eta$, every efficient stegosystem is insecure.

However, this does not imply that for a given stegosystem $\mathcal{S}$ there exists a warden W that can detect the use of $\mathcal{S}$ for *every* channel in the family $\mathcal{F} = \mathtt{PRC}_\eta$. In section 5.3 we will apply the stegosystem $\mathcal{S}_\mathcal{F}$ presented in the previous section and a slightly modified variant $\mathcal{S}_\mathcal{F}^{\hat{\omega}}$ of $\mathcal{S}_\mathcal{F}$ to the channel family $\mathcal{F} = \mathtt{PRC}_\eta$ to illustrate the properties of the measures for insecurity and detectability introduced above. Both systems are efficient and reliable, thus according to [5] must be insecure. On the other hand, the systems are not channel-universally detectable, which follows from Theorem 4 and Lemma 2, resp. Theorem 6 and Lemma 1 (assuming the existence of pseudorandom functions). Thus, both systems are *simultaneously insecure and not detectable* according to these measures. However, if one compares $\mathcal{S}_\mathcal{F}$ and $\mathcal{S}_\mathcal{F}^{\hat{\omega}}$ more thoroughly, one comes to the conclusion that the degree of insecurity/detectability should not be equal for the two systems: $\mathcal{S}_\mathcal{F}^{\hat{\omega}}$ looks far more easy to break in practice than $\mathcal{S}_\mathcal{F}$. On the contrary, determining the channel-specific detectability we will show in Theorem 5 and 6

$$\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}^{\hat{\omega}}}(t,q,\lambda) = 0 \quad \text{and} \quad \mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}}(t,q,\lambda) \geq 1-\delta$$

for a small function δ. This runs counter to our intuition regarding the strength of $\mathcal{S}_\mathcal{F}^{\hat{\omega}}$ and $\mathcal{S}_\mathcal{F}$. We therefore conclude that not only `InSec` and `UnivDetect`, but also `SpecDetect` faces serious problems in providing a reasonable measure of steganographic security.

Average detectability, on the other hand, seems to agree with our intuition. Assuming the existence of pseudorandom functions Theorem 4 and 5 imply for small functions δ and ε

$$\mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_\mathcal{F}^{\hat{\omega}}}(t,q,\lambda) \geq 1-\delta \text{ and } \mathtt{AvgDetect}_{\mathtt{PRC}_\eta,\mathcal{S}_\mathcal{F}}(t,q,\lambda) \leq \varepsilon \ .$$

These bounds also imply Theorem 1 stating that the two measures `SpecDetect` and `AvgDetect` are incomparable.

5.1 Pseudorandom Flat h-Channels

Below we recall the construction and main properties of pseudorandom flat h-channels, as given in [5]. Let Σ be an ordered alphabet of size 2^σ and $h \in [1..\sigma]$ be a chosen min-entropy. For simplicity we may assume $\Sigma = \{0, 1, \ldots, 2^\sigma - 1\}$.

First we describe a (truly random) flat h-channel specified by a probabilistic Turing machine $\mathcal{R}$ with a random tape containing an infinite random string π. For an integer tuple (σ, h, i, a, b) as input with $0 < h \leq \sigma$ and $i > 0$ and $0 \leq a \leq b < 2^\sigma$, the machine $\mathcal{R}$ does the following:
(1) it divides π into consecutive substrings of length $\zeta = 2^\sigma$ each;
(2) it identifies those substrings that have exactly 2^h ones: let y_i be the i-th such substring;
(3) it returns the number of ones in y_i between and including positions a and b (positions are counted from 0 to $2^\sigma - 1$).

Let D_i^π be the subset of Σ of cardinality 2^h that has characteristic vector y_i and let $\overrightarrow{D}^\pi := D_1^\pi \times D_2^\pi \times D_3^\pi \times \cdots$. Obviously, querying $\mathcal{R}$ with tuple (σ, h, i, a, b) returns the number of elements in $D_i^\pi \cap [a..b]$. Moreover, testing membership in D_i for an element s can be done easily by a single query to $\mathcal{R}$, namely $\chi_{D_i}(s) = \mathcal{R}(S, H, i, s, s)$.

The notation $\overrightarrow{D}^{\pi}$ will also be used for the (memoryless) channel over $D_1^{\pi} \times D_2^{\pi} \times D_3^{\pi} \times \cdots$ with uniform probability distributions, i.e. for any legal history $\mathcal{H} = s_1 s_2 \ldots s_i$ the probability distribution $\overrightarrow{D}_{\mathcal{H}}^{\pi}$ is the uniform distribution over the set D_{i+1}^{π}. Such a channel $\overrightarrow{D}^{\pi}$ is called a **truly random flat *h*-channel**.

Using techniques of Goldreich, Goldwasser, and Nussboim [7], Dedić et al. have constructed a truthful pseudorandom implementation of $\mathcal{R}$ [5]. In the construction above the truly random infinite string π is replace by a pseudorandom string π' that is generated by an appropriate pseudorandom generator from a short random seed ω of length η. This creates a **pseudorandom flat *h*-channel** $\overrightarrow{D}^{\omega}$ that is indistinguishable from the truly random flat h-channels $\overrightarrow{D}^{\pi}$. In addition, the construction allows efficient counting, membership testing and random sampling. For a seed bound η, the **family of pseudorandom flat *h*-channels** is then given by

$$\mathtt{PRC}\eta \;:=\; \{\overrightarrow{D}^{\omega} : |\omega| = \eta\} \;\; .$$

5.2 Security of Pseudorandom Functions and Channels

To analyse the quality of such a construction a security measure is needed for pseudorandom functions. The **insecurity of a family $\mathtt{PRF}\,\eta$ of pseudorandom functions** $\mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(d, t, q)$ with seed length η is defined as the advantage of an adversary to distinguish a random member of $\mathtt{PRF}\,\eta$ from a truly random function, where he has a priori information of size d about $\mathtt{PRF}\,\eta$, is t time-bounded and may ask up to q queries to a challenge oracle. Since in our setting it always holds $q \leq t$, we will skip the last parameter in PRF-InSec and write $\mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(d, t)$ to shorten the notation.

Stating the result of [5] more formally, the following has been shown.

Fact 1 *Given a family of pseudorandom functions* $\mathtt{PRF}\,\eta$*, for any* $h < \sigma$ *one can construct a family of pseudorandom flat* h*-channels* $\overrightarrow{D}^{\omega} = D_1^{\omega} \times D_2^{\omega} \times D_3^{\omega} \times \cdots$ *over an alphabet of size* 2^{σ}*, indexed by strings* ω *of length* η *such that*

1. *counting the number of elements* s*, with* $a \leq s \leq b$ *in* D_i^{ω}*, can be done in time polynomial in* η*,* σ*, and* $\log i$*, given the tuple* $(\omega, \sigma, h, i, a, b)$ *as input;*
2. *sampling and membership testing for* D_i^{ω} *can be done in time polynomial in* η*,* σ *and* $\log i$ *given the tuple* (ω, σ, h, i)*, resp.* $(\omega, \sigma, h, i, s)$ *as input;*
3. *there exists a polynomial* p *such that for every* τ *time-bounded oracle machine* $Q^{\mathcal{O}, \chi(\mathcal{O})}$ *trying to distinguish the truly random flat* h*-channel* $\overrightarrow{D}^{\pi}$ *from* $\overrightarrow{D}^{\omega}$ *using a sampling oracle* $\mathcal{O}$ *and a membership testing oracle* $\chi(\mathcal{O})$ *has only a small advantage:*

$$\Pr_{\pi}[Q^{\overrightarrow{D}^{\pi}, \chi(\overrightarrow{D}^{\pi})} = 1] - \Pr_{\omega}[Q^{\overrightarrow{D}^{\omega}, \chi(\overrightarrow{D}^{\omega})} = 1] \;\leq\; \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p(\tau, \eta)) + \frac{\tau}{2^{\eta}} \;\; . \qquad (3)$$

5.3 Upper and Lower Bounds for Detectabilities

In the following we define two stegosystems for the family of pseudorandom flat h-channels $\mathcal{F} = \mathtt{PRC}_\eta$ which may look quite similar at first glance. The first one is our generic stegosystem $\mathcal{S}_\mathcal{F}$ defined in Fig. 2 applied to this family. By Theorem 2 $\mathcal{S}_{\mathtt{PRC}\eta}$ is efficient, i.e. running in polynomial time with respect to the description size η, the length of the message n, and the size of documents σ. This follows from the properties of pseudorandom flat h-channels, namely (1) $\mathtt{PRC}_\eta$ is a family of channels such that each channel in $\mathtt{PRC}_\eta$ has description size η, (2) the distribution functions of channels in $\mathtt{PRC}_\eta$ are efficiently computable and (3) the selection of channels $\mathcal{C}_\omega$ is uniform.

Moreover, for every channel $\mathcal{C}_\omega$ in $\mathtt{PRC}_\eta$ and for every history $\mathcal{H}$ the probability distribution $\overrightarrow{D}^\omega_\mathcal{H}$ is uniform. Since the cardinality of the support of $\overrightarrow{D}^\omega_\mathcal{H}$ is a power of two, the interval encoding works perfectly. Thus, the unreliability of $\mathcal{S}_\mathcal{F}$ is zero. By Theorem 2, the security measure $\mathtt{AvgDetect}_{\mathtt{PRC}_\eta,\mathcal{S}_{\mathtt{PRC}_\eta}}$ is closely related to the distinguishability $\mathtt{Dist}_{\mathtt{PRC}_\eta}$ of flat h-channels.

The second stegosystem, denoted by $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$, works in the same manner as $\mathcal{S}_\mathcal{F}$ except the selection of the channel $\mathcal{C}_\omega$ for the interval encoding. Now this is a fixed value $\hat{\omega}$, thus a predefined part of $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$. Formally, the only difference between $\mathcal{S}_\mathcal{F}$ and $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ is that in the system $\mathcal{S}_\mathcal{F}$ both encoder and decoder use a secret (random) key ω to select a channel $\mathcal{C}_\omega$ at random while in the system $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ encoder and decoder use a predetermined value $\hat{\omega}$. According to Kerckhoffs' principle we assume that $\hat{\omega}$ is known to a warden attacking $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$, while the value ω used in $\mathcal{S}_\mathcal{F}$ remains unknown since it is part of the private key. Again $\hat{\omega}$ by $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ is *independent* of the "real" description $\omega_\mathcal{C}$ for the covertext channel $\mathcal{C}$. In $\mathcal{S}_\mathcal{F}$ Alice and Bob randomly select ω for the interval encoding, whereas in $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ they cannot even choose $\hat{\omega}$ – it is built into the system.

We have already mentioned that both stegosystems $\mathcal{S}_\mathcal{F}$ and $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ may output samples that are *not* in the support of the covertext channel $\mathcal{C}$. However, the question remains how a warden can notice this in case of a complex family of channels if his computational power is limited.

Using the lower bound on the query complexity in [5], one can deduce that in both cases the insecurity $\mathtt{InSec}_{\mathtt{PRC}_\eta,\mathcal{S}^{\hat{\omega}}_\mathcal{F}}$, resp. $\mathtt{InSec}_{\mathtt{PRC}_\eta,\mathcal{S}_\mathcal{F}}$ has to be large since the encoding complexity and unreliability are small. One can even construct quite efficient wardens that achieve a large advantage. On the other hand, it will follow from the bounds below that for all polynomial wardens the detectability $\mathtt{UnivDetect}_{\mathtt{PRC}_\eta,\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}_\eta}}$, resp. $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}_{\mathtt{PRC}_\eta}}$ is small.

Moreover, by relating the distinguishability of pseudorandom functions from random functions to the advantage of a distinguisher between random and pseudorandom flat h-channels, we can bound the distinguishability of $\mathtt{PRC}_\eta$.

Theorem 3. *Using a family of pseudorandom functions* $\mathtt{PRF}\ \eta$ *for the family* $\mathtt{PRC}\eta$ *of pseudorandom flat channels with parameters* (h, η)*, there exists a fixed polynomial* p *such that*

$$\mathtt{Dist}_{\mathtt{PRC}\eta}(t, q, \lambda) \;\leq\; 3 \cdot \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p(t, \eta)) + \frac{9\,q^2}{2^{h+1}} + \frac{3\,t}{2^\eta}\ .$$

We postpone the proof to the next section and rather continue the comparison of the two stegosystems. Combining this theorem with Theorem 2

Theorem 4. *Applied to the channel family* $\mathtt{PRC}\eta$ *of pseudorandom flat* h*-channel that is generated by a family* $\mathtt{PRF}\ \eta$ *of pseudorandom functions, the stegosystem* $\mathcal{S}_{\mathtt{PRC}\eta}$ *based on a family* $\mathtt{PRP}$ *of random permutations achieves reliability*

$$\begin{aligned}\mathtt{AvgDetect}_{\mathtt{PRC}\eta,\mathcal{S}_{\mathtt{PRC}\eta}}(t,q,\lambda) \;\le\; & 3\;\mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta,p(t,\eta)) + 2\;\mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t),\lambda/n) + \\ & (\lambda^2/n^2-\lambda/n)\cdot 2^{-n} + 9\;q^2\;2^{-(h+1)} + 3\;t\;2^{-\eta}\;.\end{aligned}$$

Thus, if pseudorandom functions and permutations with exponential small insecurity exist the average detectability of this stegosystem can also be made exponentially small. Since by Lemma 2 the relation $\mathtt{UnivDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}} \le \mathtt{AvgDetect}_{\mathcal{F},\mathcal{S}_{\mathcal{F}}}$ holds, the channel universal detectability of $\mathcal{S}_{\mathtt{PRC}\eta}$ applied to $\mathtt{PRC}\eta$ is small, too. On the other hand, the specific detectability measure gives a value arbitrarily close to 1 for $\mathcal{S}_{\mathtt{PRC}\eta}$.

Theorem 5. *There exist polynomials* p_1, p_2 *such that for* $t = p_1(\eta,\sigma,n,q)$

$$\mathtt{SpecDetect}_{\mathtt{PRC}\eta,\mathcal{S}_{\mathtt{PRC}\eta}}(t,q,nq) \;\ge\; 1 - \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta,p_2(t)) - t\;2^{-\eta} - 2^{(h-\sigma)\;q\;\ell} - (q\;\ell)^2\;2^{-h}\;.$$

Now let us perform the same estimation for the second stegosystem $\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}$. Both measures change their values drastically.

Theorem 6. *There exist a polynomial* p *such that*

$$\mathtt{SpecDetect}_{\mathtt{PRC}\eta,\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}}(t,q,\lambda) \;\le\; \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}}(p(t),q)\;.$$

Theorem 7. *For suitable polynomials* p_1, p_2 *and* $t = p_1(\eta,\sigma,n,q)$ *holds*

$$\mathtt{AvgDetect}_{\mathtt{PRC}\eta,\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}}(t,q,nq) \;\ge\; 1 - \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta,p_2(t)) + t\;2^{-\eta} + 2^{(h-\sigma)\;q\;\ell} + (q\;\ell)^2\;2^{-h}\;.$$

In practice, the system $\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}$ is easy to break when knowing its encoding channel $\mathcal{C}_{\hat{\omega}}$, whereas $\mathcal{S}_{\mathcal{F}}$ seems to be strong against any kind of attacks. As our results show these properties are reflected only by the measure *detectability on average.*

6 Formal Analysis of Distinguishability and Detectability

This section gives proofs of the main theorems.

6.1 Proof of Theorem 3

Let $\gamma(q,h) := q^2\;2^{-h}$. For the channel family $\mathcal{F} = \mathtt{PRC}_\eta$, in order to prove

$$\mathtt{Dist}_{\mathcal{F}}(t,q,\lambda) \;\le\; 3\cdot\mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta,p(t,\eta)) + \frac{9}{2}\gamma(q,h) + \frac{3\;t}{2^\eta}$$

let Q be a (t,q,λ)-distinguisher achieving maximum advantage, that means

$$\mathtt{Dist}_{\mathcal{F}}(t,q,\lambda) \;=\; \max_{Q}\Pr_{\mathcal{C}_\omega,\mathcal{C}_{\omega'}\in_R\mathtt{PRC}_\eta}[Q^{\mathcal{C}_\omega,\mathcal{C}_{\omega'}}=1] - \Pr_{\mathcal{C}_\omega\in_R\mathtt{PRC}_\eta}[Q^{\mathcal{C}_\omega,\mathcal{C}_\omega}=1]\;,$$

where $\mathcal{C}_\omega, \mathcal{C}_{\omega'}$ are random elements of $\mathtt{PRC}_\eta$ with seed ω, resp. ω'. To simplify the notation, a channel $\mathcal{C}_\omega$ and its support $\overrightarrow{D}^\omega$ will be associated. Thus instead of $\mathrm{Pr}_{\mathcal{C}_\omega, \mathcal{C}_{\omega'} \in_R \mathtt{PRC}_\eta}[Q^{\mathcal{C}_\omega, \mathcal{C}_{\omega'}} = 1]$ we simply write $\mathrm{Pr}_{\omega,\omega'}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'}} = 1]$.

Based on Q we will construct a distinguisher R working in in time $p(t, \eta)$ for some polynomial p that can detect differences between pseudorandom and truly random flat h-channels with advantage at least

$$\frac{1}{3} \cdot \Big(\mathrm{Pr}_{\omega,\omega'}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'}} = 1] - \mathrm{Pr}_{\omega}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega}} = 1] \Big) - \frac{3}{2}\, \gamma(q, h) \ .$$

Then the relation (3) in Fact 1 gives the bound stated in Theorem 3:

$$\begin{aligned}
\mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p(t,\eta)) \ &\geq \ \mathrm{Pr}_{\overrightarrow{D}}[R^{\overrightarrow{D}, \chi(\overrightarrow{D})} = 1] - \mathrm{Pr}_{\omega}[R^{\overrightarrow{D}^\omega, \chi(\overrightarrow{D}^\omega)} = 1] \ - \frac{t}{2^\eta} \\
&\geq \ \frac{1}{3} \cdot \Big| \mathrm{Pr}_{\omega,\omega'}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'}} = 1] - \mathrm{Pr}_{\omega}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega}} = 1] \Big| - \frac{3}{2}\, \gamma(q, h) - \frac{t}{2^\eta} \\
&= \ \frac{1}{3} \cdot \mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) - \frac{3}{2}\, \frac{q^2}{2^h} - \frac{t}{2^\eta} \ .
\end{aligned}$$

In the following let us abbreviate the notation by

$$\alpha_0 \ := \ \mathrm{Pr}_{\omega}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega}} = 1] \ , \quad \alpha_1 \ := \ \mathrm{Pr}_{\omega,\omega'}[Q^{\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'}} = 1] \quad \text{and} \quad \Delta := \alpha_1 - \alpha_0 \ = \ \mathtt{Dist}_{\mathcal{F}}(t, q, \lambda) \ .$$

We split the advantage Δ for differentiating between the pair of oracles $(\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega})$ and $(\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'})$ into several intermediate steps involving random h-sets and random sequences. At least one of these steps must give a significant advantage in order to gain total advantage Δ. Such a step will be exploited to design a distinguisher for the family of pseudorandom permutations generating flat h-channels.

Consider the behaviour of the distinguisher Q in cases when instead of sample sequences from oracles $(\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega})$ or $(\overrightarrow{D}^\omega, \overrightarrow{D}^{\omega'})$ sequences from some other sets, namely either from truly random flat h-sets $\overrightarrow{D}^\pi = D_1 \times D_2 \times \cdots$ or random sequences from $\overrightarrow{\Sigma} = \Sigma \times \Sigma \times \ldots$ are given. Note that in such cases Q can behave quite arbitrarily.

For $\overrightarrow{Y} = Y_1 \times Y_2 \times \ldots$ and $\overrightarrow{Z} = Z_1 \times Z_2 \times \ldots$ with $Y_i, Z_i \subseteq \Sigma$ for all i, let $Q^{\overrightarrow{Y}, \overrightarrow{Z}}$ denote Q with access to two oracles: the first one provides sequences $(s_{1,1}, s_{2,1}, \ldots, s_{\ell_1,1}), (s_{1,2}, s_{2,2}, \ldots, s_{\ell_2,2}), \ldots$ of examples with $s_{i,j} \in_R Y_j$, the second sequences $(s'_{1,1}, s'_{2,1}, \ldots, s'_{\ell'_1,1}), (s'_{1,2}, s'_{2,2}, \ldots, s'_{\ell'_2,2}), \ldots$ with $s'_{i,j} \in_R Z_j$, where all elements are chosen uniformly and independently at random.

Let us first compare the advantage Q can achieve for $(\overrightarrow{D}^\pi, \overrightarrow{D}^{\omega'})$ against $(\overrightarrow{\Sigma}, \overrightarrow{D}^\omega)$. Since the challenge oracle in both cases is a random element from the family $\mathtt{PRC}_\eta$ and unrelated to the reference oracle it suffices to bound the advantage derived from the reference oracle, that means the distance between $\overrightarrow{D}^\pi$ and $\overrightarrow{\Sigma}$. Let

$$\alpha_2 \ := \ \mathrm{Pr}_{\pi,\omega'}[Q^{\overrightarrow{D}^\pi, \overrightarrow{D}^{\omega'}} = 1] \quad \text{and} \quad \alpha_3 \ := \ \mathrm{Pr}_{\omega}[Q^{\overrightarrow{\Sigma}, \overrightarrow{D}^\omega} = 1] \ .$$

Lemma 3. *For every (t, q, λ)-distinguisher and flat h-channels holds:*

$$\alpha_2 - \alpha_3 \ = \ \mathrm{Pr}_{\pi,\omega'}[Q^{\overrightarrow{D}^\pi, \overrightarrow{D}^{\omega'}} = 1] - \mathrm{Pr}_{\omega}[Q^{\overrightarrow{\Sigma}, \overrightarrow{D}^\omega} = 1] \ \leq \ 2\, \gamma(q, h) \ .$$

Proof. Define $\psi := |\Sigma| = 2^\sigma$ and $H := 2^h$. If the reference oracle is $\overrightarrow{\Sigma}$ the elements of a sample sequence $s = s_1, s_2, \ldots, s_q$ are completely independent. The same holds for the reference oracle $\overrightarrow{D}^\pi$ for elements that come from different D_i. Thus, the best advantage will be obtained if all examples are taken from the same channel $D = D_i$ for some i. Let $X = x_1, \ldots, x_q$ be a random variable denoting the outcome of the following experiment: randomly choose $D \subseteq \Sigma$ of cardinality H and then uniformly and independently choose $x_j \in_R D$. Similarly, we define $X' = x'_1, \ldots, x'_q$ where now $x'_j \in_R \Sigma$. Then,

$$|\alpha_2 - \alpha_3| \;\leq\; \sum_{s \in \Sigma^q} \left|\Pr[X = s] - \Pr[X' = s]\right| \;.$$

Let us call a sequence X *injective* if it does not contain duplicates. Then,

$$\begin{aligned}\Pr[X = s] \;=\; &\Pr[X = s \mid X \text{ injective}] \cdot \Pr[X \text{ injective}] + \\ &\Pr[X = s \mid X \text{ not injective}] \cdot \Pr[X \text{ not injective}] \;,\end{aligned}$$

and similarly for X'. It holds

$$\Pr[X \text{ injective}] = \prod_{j=0}^{q-1} (1 - \frac{j}{H}) \;\geq\; \left(1 - \frac{q}{H}\right)^q \;\geq\; 1 - \frac{q^2}{H} \;.$$

Let s be an injective sequence. Obviously

$$\Pr[X = s \mid X \text{ not injective}] \;=\; 0 \;=\; \Pr[X' = s \mid X' \text{ not injective}] \;.$$

Furthermore, the two probabilities are also equal for X, X' injective, that is

$$\Pr[X = s \mid X \text{ injective}] \;=\; \Pr[X' = s \mid X' \text{ injective}] \;.$$

This follows from

$$\Pr[X' = s \mid X' \text{ injective}] \;=\; \frac{1}{\psi} \cdot \frac{1}{\psi - 1} \cdot \frac{1}{\psi - 2} \cdots \frac{1}{\psi - q + 1} \;.$$

and the calculation

$$\Pr[X = s \mid X \text{ injective}] \;=\; \frac{\binom{\psi - q}{H - q}}{\binom{\psi}{H}} \cdot \frac{1}{H} \cdot \frac{1}{H - 1} \cdot \ldots \frac{1}{H - q + 1} \;=\; \frac{1}{\psi} \cdot \frac{1}{\psi - 1} \cdot \frac{1}{\psi - 2} \cdots \frac{1}{\psi - q + 1} \;.$$

Since the domain of X is smaller than the one of X' it hold $\Pr[X \text{ not injective}] \ \geq \ \Pr[X' \text{ not injective}]$. This implies

$$\begin{aligned}
&\sum_{s\in\Sigma^q} \big|\Pr[X=s] - \Pr[X'=s]\big| \\
&= \sum_{s \text{ injective}} \big|\Pr[X=s \mid X \text{ inj.}]\cdot\Pr[X \text{ inj.}] - \Pr[X'=s \mid X' \text{ inj.}]\cdot\Pr[X' \text{ inj.}]\big| \\
&+ \sum_{s \text{ not injective}} \big|\Pr[X=s \mid X \text{ not inj.}]\cdot\Pr[X \text{ not inj.}] - \Pr[X'=s \mid X' \text{ not inj.}]\cdot\Pr[X' \text{ not inj.}]\big| \\
&\leq \sum_{s \text{ injective}} \Pr[X=s \mid X \text{ inj.}]\cdot\big|\Pr[X \text{ inj.}] - \Pr[X' \text{ inj.}]\big| \ + \ \Pr[X \text{ not injective}] \\
&\leq \big|\Pr[X' \text{ injective}] - \Pr[X \text{ injective}]\big| \ + \ \Pr[X \text{ not injective}] \\
&\leq \left|1 - \left(1 - \frac{q^2}{H}\right)\right| \ + \ \frac{q^2}{H} \ = \ 2\,\frac{q^2}{2^h} \ = \ 2\,\gamma(q,h) \ .
\end{aligned}$$

□

A similar estimation bounds the advantage for $(\overrightarrow{D}^\pi, \overrightarrow{D}^\pi)$ against $(\overrightarrow{\Sigma}, \overrightarrow{D}^\pi)$. Let

$$\alpha_4 \ := \ \Pr_\pi[Q^{\overrightarrow{D}^\pi,\overrightarrow{D}^\pi} = 1] \quad \text{and} \quad \alpha_5 \ := \ \Pr_\pi[Q^{\overrightarrow{\Sigma},\overrightarrow{D}^\pi} = 1] \ .$$

Lemma 4. *For every (t,q,λ)-distinguisher and flat h-channels holds:*

$$\alpha_4 - \alpha_5 \ = \ \Pr_\pi[Q^{\overrightarrow{D}^\pi,\overrightarrow{D}^\pi} = 1] - \Pr_\pi[Q^{\overrightarrow{\Sigma},\overrightarrow{D}^\pi} = 1] \ \leq \ 2\,\gamma(q,h) \ .$$

Now we design a distinguisher R with advantage

$$\Pr_\omega[R^{\overrightarrow{D}^\omega,\,\chi(\overrightarrow{D}^\omega)} = 1] - \Pr_\pi[R^{\overrightarrow{D}^\pi,\,\chi(\overrightarrow{D}^\pi)} = 1] \ \geq \ \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h) \ .$$

R will not make use of membership queries at all, thus we can simplify this advantage to

$$\Pr_\omega[R^{\overrightarrow{D}^\omega} = 1] - \Pr_\pi[R^{\overrightarrow{D}^\pi} = 1] \ \geq \ \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h) \ . \tag{4}$$

1. If $\alpha_0 - \alpha_4 \geq \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)$, then $R^{\overrightarrow{X}}$ with $\overrightarrow{X}$ either $\overrightarrow{D}^\omega$ or $\overrightarrow{D}^\pi$ simulates $Q^{\overrightarrow{D}^\omega,\overrightarrow{D}^\omega}$, resp. $Q^{\overrightarrow{D}^\pi,\overrightarrow{D}^\pi}$ as follows. Whenever Q requires an example of length ℓ from either oracle, R obtains an example sequence $(s_1, s_2, \ldots, s_\ell)$, with $s_i \in X_i$ from $\overrightarrow{X}$, and provides this sequence to Q. Finally, R outputs the value that Q has returned. This means

   $$\Pr_\omega[R^{\overrightarrow{D}^\omega} = 1] = \Pr_\omega[Q^{\overrightarrow{D}^\omega,\overrightarrow{D}^\omega} = 1] \quad \text{and} \quad \Pr_\pi[R^{\overrightarrow{D}^\pi} = 1] = \Pr_\pi[Q^{\overrightarrow{D}^\pi,\overrightarrow{D}^\pi} = 1] \ ,$$

   which implies

   $$\Pr_\pi[R^{\overrightarrow{D}^\pi} = 1] - \Pr_\omega[R^{\overrightarrow{D}^\omega} = 1] \ = \ \alpha_0 - \alpha_4 \ \geq \ \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h) \ .$$

2. If $\alpha_1 - \alpha_2 \geq \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)$, then $R^{\overrightarrow{X}}$ with $\overrightarrow{X}$ either $\overrightarrow{D}^{\omega}$ or $\overrightarrow{D}^{\pi}$ simulates $Q^{\overrightarrow{X},\overrightarrow{D}^{\omega'}}$ by choosing ω' randomly. Whenever Q requires an example of length ℓ from the first oracle, R, similarly as in the previous case, obtains an example sequence $(s_1, s_2, \ldots, s_\ell)$ from $X_1 \times X_2 \times \ldots \times X_\ell$ and provides it to Q. If Q needs an example from the second oracle, then R uses ω' to simulate $\overrightarrow{D}^{\omega'}$ and provides $(s_1, s_2, \ldots, s_\ell)$ to Q. As before, R outputs the same value as Q. It holds

$$\Pr_{\omega}[R^{\overrightarrow{D}^{\omega}} = 1] = \Pr_{\omega,\omega'}[Q^{\overrightarrow{D}^{\omega},\overrightarrow{D}^{\omega'}} = 1] \quad \text{and} \quad \Pr_{\pi}[R^{\overrightarrow{D}^{\pi}} = 1] = \Pr_{\pi,\omega'}[Q^{\overrightarrow{D}^{\pi},\overrightarrow{D}^{\omega'}} = 1]\ ,$$

$$\text{thus} \quad \Pr_{\pi}[R^{\overrightarrow{D}^{\pi}} = 1] - \Pr_{\omega}[R^{\overrightarrow{D}^{\omega}} = 1] \;=\; \alpha_1 - \alpha_2 \;\geq\; \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)\ .$$

3. If $\alpha_3 - \alpha_5 \geq \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)$, then $R^{\overrightarrow{X}}$ with $\overrightarrow{X}$ either $\overrightarrow{D}^{\omega}$ or $\overrightarrow{D}^{\pi}$ simulates $Q^{\overrightarrow{\Sigma},\overrightarrow{X}}$. During the simulation, whenever Q requires an example of length ℓ from the first oracle, R chooses uniformly and independently at random elements $s_i \in_R \Sigma$ for $i = 1, \ldots, \ell$ and provides $(s_1, s_2, \ldots, s_\ell)$ to Q. For an example sequence from the second oracle R passes a sequence $(s_1, s_2, \ldots, s_\ell)$ from $X_1 \times X_2 \times \ldots \times X_\ell$ to Q and outputs what Q has returned. It follows

$$\Pr_{\omega}[R^{\overrightarrow{D}^{\omega}} = 1] = \Pr_{\omega}[Q^{\overrightarrow{\Sigma},\overrightarrow{D}^{\omega}} = 1] \quad \text{and} \quad \Pr_{\pi}[R^{\overrightarrow{D}^{\pi}} = 1] = \Pr_{\pi}[Q^{\overrightarrow{\Sigma},\overrightarrow{D}^{\pi}} = 1]\ ,\ \text{thus}$$

$$\Pr_{\overrightarrow{D}}[R^{\overrightarrow{D}} = 1] - \Pr_{\omega}[R^{\overrightarrow{D}^{\omega}} = 1] \;=\; \alpha_3 - \alpha_5 \;\geq\; \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)\ .$$

Thus, in each case we are able to provide a distinguisher that achieves the advantage stated in (4). Finally we show that at least one of these cases has to occur. Assume to the contrary that this does no hold, that means

$$\max\{\alpha_0 - \alpha_4,\ \alpha_1 - \alpha_2,\ \alpha_3 - \alpha_5\} \;<\; \frac{\Delta}{3} - \frac{3}{2}\,\gamma(q,h)\ .$$

Then, using Lemma 3 to bound the difference between α_2 and α_3 one can deduce

$$\alpha_1 - \alpha_5 = (\alpha_1 - \alpha_2) + (\alpha_2 - \alpha_3) + (\alpha_3 - \alpha_5) \;<\; \frac{2\,\Delta}{3} - 3\,\gamma(q,h) + 2\,\gamma(q,h) \;=\; \frac{2\,\Delta}{3} - \;\gamma(q,h)\ .$$

On the other hand, applying Lemma 4 we get

$$\alpha_0 - \alpha_5 \;=\; (\alpha_0 - \alpha_4) + (\alpha_4 - \alpha_5) \;<\; \frac{\Delta}{3} + \frac{1}{2}\,\gamma(q,h)\ .$$

Since $\Delta = \alpha_1 - \alpha_0$ this gives the contradiction

$$\Delta \;=\; \alpha_1 - \alpha_0 \;=\; (\alpha_1 - \alpha_5) + (\alpha_0 - \alpha_5) \;<\; \Delta - \frac{1}{2}\,\gamma(q,h)\ .$$

□

6.2 Proof of Theorem 5

We have to show for suitable polynomials p_1 and p_2 and $t = p_1(\eta, \sigma, n, q)$

$$\mathtt{SpecDetect}_{\mathtt{PRC}\eta, \mathcal{S}_{\mathtt{PRC}\eta}}(t, q, nq) \;\geq\; 1 - \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p_2(t)) - t\; 2^{-\eta} - 2^{(h-\sigma)\; q\; \ell} - (q\;\ell)^2\; 2^{-h}\;.$$

This means for an arbitrary, but fixed channel $\mathcal{C} \in \mathtt{PRC}\eta$ we have to construct a (t, q, nq)-warden W with advantage $\mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C}, \mathcal{S}_{\mathcal{F}}}(W) \geq 1 - \delta'$ where

$$\delta' \;:=\; \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p_2(t)) + t\; 2^{-\eta} + 2^{(h-\sigma)\; q\; \ell} + (q\;\ell)^2\; 2^{-h}\;.$$

Let $\omega_{\mathcal{C}}$ be the seed of $\mathcal{C}$. W makes q queries to the challenge oracle CH which is either $\mathcal{S}_{\mathcal{F}}$-$\mathbf{Encode}^{\mathcal{C}_\omega}(K, M, \mathcal{H})$ or $OC(M, \mathcal{H})$. To achieve the required advantage, he chooses an arbitrary message $M \in \{0,1\}^n$ and queries the challenge oracle with M and the empty history $\mathcal{H}$.

For the analysis it will be important that the return $s_1, \ldots, s_{2\ell}$ of $\mathcal{S}_{\mathcal{F}}$-$\mathbf{Encode}^{\mathcal{C}_\omega}(K, M, \mathcal{H})$ has the following property: the first ℓ elements $s_1, \ldots, s_\ell$ are random elements from the support $D_1^{\omega_{\mathcal{C}}} \times \ldots \times D_\ell^{\omega_{\mathcal{C}}}$. Only this part of each sample sequence will be used by W. Formally, using $\omega_{\mathcal{C}}$ as a predefined parameter he executes the following steps:

1. choose an arbitrary message $M \in \{0,1\}^n$;
2. for $i = 1, \ldots, q$ do:
   (a) query the challenge oracle CH with M and the empty history $\mathcal{H}$; let $S_i = s_{i,1}, \ldots, s_{i,2\ell}$ be the output of CH;
   (b) use the membership test for $\omega_{\mathcal{C}}$ on the first ℓ elements $s_{i,1}, \ldots, s_{i,\ell}$;
   (c) if there exists some $s_{i,j} \notin \overrightarrow{D_j}^{\omega_{\mathcal{C}}}$, then terminate and output 1 (for "stego");
3. finally output 0 (for "no stego") if termination has not occurred before.

W takes time polynomial in η, σ, n and q and makes q queries of total length $\lambda = qn$. By the construction of the stegosystem $\mathcal{S}_{\mathcal{F}}$ we know that the probability distribution of the strings $s_{i,1}, \ldots, s_{i,\ell}$ is exactly the same as in the channel $\mathcal{C}_\omega$, where ω is the private key of Alice used for the encoding in $\mathcal{S}_{\mathcal{F}}$. W will always correctly output 0 if it sees original samples from $\mathcal{C}$, therefore $\Pr[W^{\mathcal{C},\mathcal{C}} = 1] = 0$. Thus, the advantage of W is given by

$$\mathtt{Adv}^{\mathsf{cha}}_{\mathtt{PRC}\eta, \mathcal{S}_{\mathtt{PRC}\eta}}(W) \;=\; \Pr_\omega[W^{\mathcal{C},\mathcal{C}_\omega} = 1] - \Pr[W^{\mathcal{C},\mathcal{C}} = 1] \;=\; 1 - \Pr_\omega[W^{\mathcal{C},\mathcal{C}_\omega} = 0]\;,$$

Again we associate a channel and its support, thus it remains to show

Lemma 5.

$$\Pr_\omega[W^{\mathcal{C},\mathcal{C}_\omega} = 0] \;=\; \Pr_\omega[W^{\overrightarrow{D}^{\omega_{\mathcal{C}}}, \overrightarrow{D}^{\omega}} = 0] \;\leq\; \delta'\;.$$

Proof. By the construction of W it holds

$$\Pr_\omega[W^{\overrightarrow{D}^{\omega_{\mathcal{C}}}, \overrightarrow{D}^{\omega}} = 0] \;=\; \Pr_{\omega;\; S_1, \ldots, S_q \in_R \overrightarrow{D}^{\omega}}[S_1, \ldots, S_q \in \overrightarrow{D}^{\omega_{\mathcal{C}}}]\;.$$

If this probability were high one could distinguish easily between a truly random flat h-channel $\overrightarrow{D}^{\pi}$ and a pseudorandom $\overrightarrow{D}^{\omega}$ because $\Pr_{\overrightarrow{D}^{\pi};\; S_1, \ldots, S_q \in_R \overrightarrow{D}^{\pi}}[S_1, \ldots, S_q \in \overrightarrow{D}^{\omega_{\mathcal{C}}}]$ is negligible:

Lemma 6. $\Pr_{\pi;\, S_1,\ldots,S_q \in_R \overrightarrow{D}^\pi}[S_1,\ldots,S_q \in \overrightarrow{D}^{\omega\mathcal{C}}] \;\leq\; 2^{(h-\sigma)\; q\; \ell} + (q\;\ell)^2\; 2^{-h}$.

Proof. If the collection of the first ℓ elements of all S_i does not contain any duplicates the probability that a single element $s_{i,j}$ with $i \in [1..q]$ and $j \in [1..\ell]$ belongs to $\overrightarrow{D_j}^{\omega\mathcal{C}}$ is simply $2^{h-\sigma}$. The probability that no duplicates occur is at most

$$\prod_{u=0}^{q\ell-1}\left(1-\frac{u}{2^h}\right) \;\geq\; \left(1-\frac{q\;\ell}{2^h}\right)^{q\;\ell} \;\geq\; \exp\left(-\frac{(q\;\ell)^2}{2^h}\right) \;\geq\; 1-\frac{(q\;\ell)^2}{2^h}\ .$$

□

Using this observation we construct a distinguisher Q which simulates W in order to distinguish the truly random flat h-channels from pseudorandom ones. Then we can use the upper bound (3) on the advantage of such a distinguisher. Q only queries the sample oracle $\mathcal{O}$ that can be either a truly random $\overrightarrow{D}^\pi$ or a pseudorandom $\overrightarrow{D}^\omega$. It does not use the membership oracle $\chi(\mathcal{O})$ at all. Q asks the oracle $\mathcal{O}$ to give q samples $S'_1,\ldots,S'_q$, each of length 2ℓ. Next Q simulates q iterations of the warden W skipping W's queries to the challenge oracle CH and using $S'_1,\ldots,S'_q$ instead of $S_1,\ldots,S_q$. If the warden outputs 0 for "no stego", the distinguisher Q outputs 1 for "$\overrightarrow{D}^\omega$"; if the warden outputs 1, the distinguisher Q outputs 0 for "$\overrightarrow{D}^\pi$". This way, we can bound the advantage as follows:

$$\begin{aligned}\mathtt{Adv}(Q) \;&=\; \Pr_\pi[Q^{\overrightarrow{D}^\pi}=1] - \Pr_\omega[Q^{\overrightarrow{D}^\omega}=1]\\ &=\; \Pr_\pi[W^{\overrightarrow{D}^{\omega\mathcal{C}},\overrightarrow{D}^\pi}=0] - \Pr_\omega[W^{\overrightarrow{D}^{\omega\mathcal{C}},\overrightarrow{D}^\omega}=0]\ .\end{aligned}$$

Using Lemma 6 and the bound (3) on the advantage we can conclude for suitable polynomials p_1 and p_2 and $t = p_1(\eta,\sigma,n,q)$:

$$\begin{aligned}\Pr_\omega[W^{\overrightarrow{D}^{\omega\mathcal{C}},\overrightarrow{D}^\omega}=0] \;&=\; Prob_\pi[W^{\overrightarrow{D}^{\omega\mathcal{C}},\overrightarrow{D}^\pi}=0] + \mathtt{Adv}(Q)\\ &\leq\; 2^{(h-\sigma)\; q\; \ell} + (q\;\ell)^2\; 2^{-h} + \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p_2(t)) + t\; 2^{-\eta}\ .\end{aligned}$$

□

6.3 Proof of Theorem 6

To bound $\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}}(t,q,\lambda)$ from above consider $\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}$ applied to the covertext channel $\mathcal{C}_{\hat{\omega}}$. In this situation IntervalEncode$(\omega,\mathcal{H},\rho)$ produces only stegotexts that belong to the support of $\mathcal{C}_{\hat{\omega}}$. If all u_j of the procedure $\mathcal{S}_{\mathcal{F}}$-Encode in Fig. 2 were truly random the same would hold for the ρ_j. Then the distribution generated by the procedure IntervalEncode would be identical to the one of the real channel $\mathcal{C}_{\hat{\omega}}$. This, however, only holds for the first half of the u_j generated by the random string T_0. The second half is functionally dependent on $T_0 \oplus M$ and the random permutation chosen by Alice's secret key K_1.

Consider a (t,q,λ)-warden $\hat{W}$ that achieves an optimal advantage for the covertext channel $\mathcal{C}_{\hat{\omega}}$. Because the channel is fixed the adversary is assumed to know $\hat{\omega}$. Thus, he can easily generate samples of $\mathcal{C}_{\hat{\omega}}$ by himself and does not need the reference oracle at all. His advantage can therefore simply be estimated by

$$\Pr_{K_1\in_R\{0,1\}^\kappa}[\hat{W}^{F(K_1,\cdot,\cdot)}=1] - \Pr[\hat{W}^{OC(\cdot,\cdot)}=1]\ ,$$

where $F(K_1, \mathcal{H}, M)$ denotes the result produced by $\mathcal{S}_{\mathcal{F}}$-Encode on input $(\hat{\omega}||K_1, \mathcal{H}, M)$. Using $\hat{W}$ we construct a distinguisher $\hat{D}$ for the family of pseudorandom permutations $\mathtt{PRP}_{K_1}$. The distinguisher simulates $\hat{W}$ as follows. Whenever $\hat{W}$ asks his challenge oracle to either get an element of $F(K_1, \mathcal{H}, M)$ or a random element of $\mathcal{C}_{\hat{\omega}}$, $\hat{D}$ chooses a random $T_0 \in \{0,1\}^n$ and provides $T_0 \oplus M$ to his challenge oracle. This will answer either with a string $T_1 = \mathtt{PRP}_{K_1}(T_0 \oplus M)$ or a random string of length n. Then $\hat{D}$ executes the for-loop of $\mathcal{S}_{\mathcal{F}}$-Encode to compute the output $s_1, \ldots, s_{2\ell}$. This output is passed on to $\hat{W}$. Finally, $\hat{D}$ makes the same decision as $\hat{W}$.

If the challenge oracle of $\hat{D}$ is $\mathtt{PRP}_{K_1}$ then by construction $\hat{D}$ produces the same distribution for $\hat{W}$ as his challenge oracle, thus

$$\Pr_{K_1 \in_R \{0,1\}^\kappa}[\hat{D}^{\mathtt{PRP}_{K_1}(\cdot)} = 1] \;=\; \Pr_{K_1 \in_R \{0,1\}^\kappa}[\hat{W}^{F(K_1,\cdot,\cdot)} = 1] \;.$$

If on the other hand $\hat{D}$ gets random strings T_1 from his oracle then the outputs s_i are completely random strings of $\mathcal{C}_{\hat{\omega}}$. Thus,

$$\Pr_{P \in_R PERM(l)}[\hat{D}^{P(\cdot)} = 1] \;=\; \Pr[\hat{W}^{OC(\cdot,\cdot)} = 1] \;.$$

Thus, $\hat{D}$ achieves the same advantage as $\hat{W}$. Furthermore, he asks exactly the same number of questions as $\hat{W}$. Let $p(t)$ be an upper time bound for $\hat{D}$ to simulate a t-time bounded warden $\hat{W}$ plus the time to compute the outputs of $\mathcal{S}_{\mathcal{F}}$-Encode. Then we can conclude

$$\begin{aligned}
\mathtt{SpecDetect}_{\mathcal{F},\mathcal{S}^{\hat{\omega}}_{\mathcal{F}}}(t,q,\lambda) \;&=\; \min_{\mathcal{C}\in\mathcal{F}} \max_{W} \; \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C},\mathcal{S}}(W) \\
&\leq\; \max_{W} \; \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C}_{\hat{\omega}},\mathcal{S}}(W) \;=\; \mathtt{Adv}^{\mathsf{cha}}_{\mathcal{C}_{\hat{\omega}},\mathcal{S}}(\hat{W}) \;=\; \mathtt{PRP\text{-}Adv}_{\mathtt{PRP}_{K_1}}(\hat{D}) \\
&\leq\; \max_{D} \mathtt{PRP\text{-}Adv}_{\mathtt{PRP}_{K_1}}(D) \;=\; \mathtt{PRP\text{-}InSec}_{\mathtt{PRP}_{K_1}}(p(t),q) \;.
\end{aligned}$$

□

6.4 Proof of Theorem 7

Concerning the detectability on average, for arbitrary q we have to show

$$\mathtt{AvgDetect}_{\mathtt{PRC}\eta,\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}}(t,n,q) \;\geq\; 1 - \mathtt{PRF\text{-}InSec}_{\mathtt{PRF}\,\eta}(\eta, p_2(t)) + t\; 2^{-\eta} + 2^{(h-\sigma)\;q\;\ell} + (q\;\ell)^2\; 2^{-h}$$

with $t = p_1(\eta, \sigma, n, q)$. To get a warden $\hat{W}$ with a large advantage we can make a similar construction as in the proof of Theorem 5 except that the warden now knows the seed $\hat{\omega}$ used by Alice, while the real covertext channel remains unknown to him. In the proof of Theorem 5 just the opposite situation occurs.

$\hat{W}$ makes q queries to the challenge oracle CH, which is either $\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}$-**Encode**$(K, M, \mathcal{H})$ or $OC(M, \mathcal{H})$. In detail, he performs the following steps.

1. Chose an arbitrary message $M \in \{0,1\}^n$;
2. for $i = 1, \ldots, q$ do
   (a) query the challenge oracle CH with M and the empty history $\mathcal{H}$; let $s_{i,1}, \ldots, s_{i,2\ell}$ be the output of CH;
   (b) use the membership test for $\hat{\omega}$ on the first ℓ elements $s_{i,1}, \ldots, s_{i,\ell}$;
   (c) if there exists some $s_{i,j} \notin \overrightarrow{D_j}^{\hat{\omega}}$, then terminate and output 0 (for "no stego");

3. finally output 1 (for "stego") if it has not terminated before.

In the stego case $\hat{W}$ will always decide 1 because all $s_{i,j}$ belong to $\overrightarrow{D_j^{\hat{\omega}}}$ by construction. The only wrong decision can occur in the nonstego case when all covertext samples by chance fall in the support of $\mathcal{C}_{\hat{\omega}}$ and $\hat{W}$ decides 1. Notice that $\hat{W}$ faces a dual situation compared to W constructed in the proof of Theorem 5. $\hat{W}$ knows the coding channel, but not the covertext channel, whereas W knows the covertext channel, but not the coding channel chosen by Alice, and the unknown channel in both situations is uniformly distributed. W only makes a wrong decision in the stego case if all stegotexts by chance fall into the support of the channel C. Therefore, the probability that $\hat{W}$ decides 1 in case of nonstego is identical to W deciding 0 in case of stego:

$$\Pr_\omega[\hat{W}^{\mathcal{C}_\omega,\mathcal{C}_\omega} = 1] \;=\; \Pr_\omega[W^{\mathcal{C},\mathcal{C}_\omega} = 0] \;.$$

By Lemma 5 this is bounded by δ'. Hence, the advantage can be estimated by

$$\begin{aligned}
\mathtt{Adv}^{\mathsf{cha}}_{\mathtt{PRC}\eta,\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}}(\hat{W}) \;&=\; \Pr_{\mathcal{C}\in_R\mathtt{PRC}\eta,K}[\hat{W}^{\mathcal{C},\mathcal{S}^{\hat{\omega}}_{\mathtt{PRC}\eta}\text{-Encode}^{\mathcal{C}}(K,\cdot,\cdot)} = 1] - \Pr_{\mathcal{C}\in_R\mathtt{PRC}\eta}[\hat{W}^{\mathcal{C},OC(\cdot,\cdot)} = 1] \\
&=\; \Pr_\omega[\hat{W}^{\mathcal{C}_\omega,\mathcal{C}_{\hat{\omega}}} = 1] - \Pr_\omega[\hat{W}^{\mathcal{C}_\omega,\mathcal{C}_\omega} = 1] \\
&=\; 1 - \Pr_\omega[\hat{W}^{\mathcal{C}_\omega,\mathcal{C}_\omega} = 1] \\
&\geq\; 1 - \delta' \;.
\end{aligned}$$

□

7 Conclusions and Future Work

A meaningful security measure for stegosystems should account for universality with respect to covertext channels as well as detection since typically neither the stegoencoder nor the stegodetector have precise knowledge about the channel. We propose to replace the notion of *insecurity* by *detectability*. Comparing three possible variants *specific detectability*, *universal detectability* and *detectability on average* that model different preconditions of the battle between the stegoencoder and the detector we have shown that only the last one can provide meaningful results. Furthermore, the detectability of a stegosystem is closely related to the difficulty to learn the covertext distribution. We have proven a tight analytical relationship between these tasks.

For a particular family of covertext channels, the pseudorandom flat h-channels, two stegosystems $\mathcal{S}_\mathcal{F}$ and $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ have been constructed with the properties: (1) both are *insecure*, (2) $\mathcal{S}_\mathcal{F}$ is not *universally detectable*, but *specifically detectable* and (3) $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ is neither *universally detectable* nor *specifically detectable*. However, low universal detectability is easy to achieve since $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ only needs to be secure for a single channel. Low specific detectability can be a misleading property, too: in practice, $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ is much easier to detect than $\mathcal{S}_\mathcal{F}$. Therefore, we settle on *detectability on average* as a "reasonable" measure for security.

It is shown that $\mathcal{S}^{\hat{\omega}}_\mathcal{F}$ is highly *detectable on average*, whereas $\mathcal{S}_\mathcal{F}$ is just the extreme opposite. This makes $\mathcal{S}_\mathcal{F}$ an interesting candidate for a stegosystem with desirable properties: it is reliable, efficient – in contrast to systems based on rejection sampling [9, 5], its sampling complexity is linear, not exponential – and still provides a good amount of security, since on average the chances that an adversary running in polynomial time can detect it are extremely low. We have given an analytical proof for this property relating the advantage of an adversary to the chance of distinguishing

different channels of the family and the chance to recognize pseudorandom permutations that are used for the construction of the stegotexts. Thus, similar to many primitives in cryptography secure steganography depends on the existence of secure pseudorandom functions. This issue needs further clarification in a strict analytical sense.

Furthermore, we propose to design and investigate other stegosystems in this setting. Can one get similar results if the pseudorandom functions used in the constructions here are replaced by cryptographic functions?

References

1. von Ahn, L., Hopper, N.J.: Public-key steganography. In: Proc. Eurocrypt 2004, Vol. 3027 of LNCS, Springer (2004), 323–341.
2. Backes, M., Cachin, C.: Public-key steganography with active attacks. In: Proc. Theory of Cryptography Conference (TCC 2005), Vol. 3378 of LNCS, Springer (2005), 210–226.
3. Bellare, M., Desai, A., Jokipii, E., Rogaway, F.: A concrete security treatment of symmetric encryption. In: Proc. 38. Symp. on Foundations of Computer Science (FOCS 1997), IEEE Computer Society (1997), 394-403; a full paper available under `www-cse.ucsd.edu/~adesai/papers/pubs.html#BDJR97` .
4. Cachin, C.: An information-theoretic model for steganography. Information and Computation 192(1) (2004), 41–56.
5. Dedić, N., Itkis, G., Reyzin, L., Russell, S.: Upper and lower bounds on black-box steganography. Journal of Cryptology 22(3) (2009), 365–394.
6. Fridrich, J., Goljan, M., Hogea, D.: Steganalysis of JPEG images: breaking the F5 algorithm. In: Proc. 8. Int. Workshop on Information Hiding (IH 2006), Vol. 2578 of LNCS, Springer (2003) 310-323.
7. Goldreich, O., Goldwasser, S., Nussboim, A.: On the implementation of huge random objects. In: 44. Symp. on Foundations of Computer Science (FOCS 2003), IEEE Computer Society (2003), 68–79.
8. Hopper, N.: On steganographic chosen covertext security. In: Proc. 32. ICALP 2005, Vol. 3580 of LNCS, Springer (2005) 311–323.
9. Hopper, N.J., Langford, J., von Ahn, L.: Provably secure steganography. In: Proc. Advances in Cryptology (CRYPTO 2002), Vol. 2442 of LNCS, Springer (2002), 77–92.
10. Ker, A., Bas, P., Böhme, R., Cogranne, R., Craver, S., iller, T., Fridrich, J., Pevnỳ, T.: Moving steganography and steganalysis from the laboratory into the real world. In: Proc. 1. Workshop on Information Hiding and Multimedia Security, ACM (2013), 45–58.
11. Le, T.V., Kurosawa, K.: Bandwidth optimal steganography secure against adaptive chosen stegotext attacks. In: Proc. 8. Information Hiding Workshop (IH 2006), Vol. 4437 of LNCS, Springer (2007), 297–313.
12. Liśkiewicz, M., Reischuk, R., Wölfel, U.: Grey-box steganography. Theoretical Computer Science 505 (2013), 27–41.
13. Lysyanskaya, A., Meyerovich, M.: Provably secure steganography with imperfect sampling. In: Proc. 9. Int. Conference on Theory and Practice in Public-Key Cryptography (PKC 2006), Vol. 3958 of LNCS, Springer (2006), 123–139.
14. Rogaway, P.: Nonce-based symmetric encryption. In: Proc. 11. Int. Workshop on Fast Software Encryption (FSE 2004), Vol. 3017 of LNCS, Springer (2004), 348–359.
15. Westfeld W.: F5 – a steganographic algorithm. In: Proc. 8. Int. Workshop on Information Hiding (IH 2006), Vol. 2578 of LNCS, Springer (2003), 289–302.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de