More Efficient Provably Secure Steganography

Leonid Reyzin and Scott Russell*
Department of Computer Science
Boston University
111 Cummington St.
Boston, MA 02215, USA
{reyzin,srussell}@bu.edu

May 15, 2003

Abstract

Steganography is the science of hiding the very *presence* of a secret message within a public communication channel. In Crypto 2002, Hopper, Langford, and von Ahn proposed the first complexity-theoretic definition and constructions of stegosystems. They later pointed out and corrected a flaw in one of their basic constructions. The correction, unfortunately, introduced a need for expensive error-correcting codes.

We obtain a more efficient stegosystem by first analyzing the severity of the flaw in their original construction. As a result, for high-entropy channels, our construction is at least 5 times more efficient (in terms of rate) than their corrected version, and requires no computationally intensive error correcting.

1 Introduction

1.1 Background

Steganography's goal is to conceal the presence of a secret message within an innocuous-looking communication. In other words, steganography consists of hiding a secret *hiddentext* message within a public *covertext* to obtain a *stegotext* in such a way that any observer (except, of course, the intended recipient) is unable to distinguish between a covertext *with* a hiddentext and one *without*. In CRYPTO 2002, Hopper, Langford and von Ahn [6] offer the first rigorous complexity-theoretic formulation of steganography. They formally define *steganographic secrecy* of a stegosystem as the inability of a polynomial-time adversary to distinguish between observed distributions of unaltered covertexts and stegotexts. This brings steganography into the realm of cryptography, unlike many previous works, which tended to be information-theoretic in perspective (see, e.g., [2] and other references in [6]).

The model assumes that the two communicating parties have some underlying distribution D of covertexts that the adversary expects to see. All parties are allowed to draw from D; the game for the sender is to alter D imperceptibly for the adversary, while transmitting a meaningful hiddentext message to the receiver. A *universal* stegosystem is one that works for *any* underlying covertext distribution D with sufficient entropy, accessing it merely as an oracle.

In addition to providing a model, the authors of [6] also present a number of constructions satisfying the definition. The most elementary one (called "Construction 1"), on which other constructions rely heavily,

*This research was facilitated in part by a National Physical Science Consortium Fellowship and by stipend support from the National Security Agency.

contains a subtle but crucial security flaw subsequently corrected by the authors in [5]. The corrected Construction 1 is universal.

Unfortunately, the price for universality of the corrected Construction 1 is very high. If D is used as a black box with unknown entropy (beyond a certain required minimum), one needs to send 22 elements of D to reliably transmit a single bit of hiddentext. Moreover, the limitation of Construction 1 is inherent: even if one gives up universality and allows the stego-encoder and -decoder knowledge of the entropy of D, it will always require at least 5 elements of D to reliably transmit a single bit of hiddentext, no matter how "good" D may be.

The reason for such high cost is the high probability of incorrectly decoding an encoded bit. To provide reliability, therefore, one has to encode the hiddentext in an error-correcting code and then stego-encode the codewords.[1] The high rate of error in stego-encoding (between $1/4$ and $3/8$, depending on D) provides an easy upper bound on the rate of the error-correcting code used, and thus a lower bound on the stretch factor, which must be between $1/(1 - H(3/8)) \approx 22$ and $1/(1 - H(1/4)) \approx 5$.

1.2 Our Contribution

We propose a new stegosystem that is able to take better advantage of the entropy inherent in D. In particular, for distributions with sufficiently high minimum entropy, we can reliably transmit one bit of hiddentext per one element of D without the need for any error-correcting codes. Our construction is thus more efficient with respect to both the running time and the message length.

Our construction is not universal only in the sense that the stego-encoder and stego-decoder need to know (a lower bound on) the minimum entropy of D (equivalently, an upper bound on the maximum probability of an element of D). It does not rely on other properties of D in any way, and will work for any D with nonzero minimum entropy. Hence, using parameter values corresponding to the worst possible D it can be made universal. As demonstrated in Section 6, for reasonable security parameters, our stegotext message length becomes shorter than the revised version of Construction 1 presented in [5] whenever the minimum entropy of D is at least 21.

The technical approach of our work is to bound the security flaw in the original (uncorrected) Construction 1 and to demonstrate that it is irrelevant provided D has sufficiently high min-entropy. We then show that if D does not have such min-entropy, it can be fixed by using D^n (for a small n) instead of D. The main technical difficulties arise in bounding the security flaw as a function of the covertext distribution D and n.

We present the background definitions and construction of [6] in Section 2. We then analyze and bound the flaw of Construction 1 in Section 3. Our new stegosystem is presented in Section 4, analyzed in Section 5 and compared with the corrected Construction 1 of [5] in Section 6.

2 Background: Work of Hopper, Langford, and von Ahn

2.1 Definitions

We reiterate the main definitions and notational conventions from [6] which we utilize herein. Many of these are taken nearly verbatim from the original work.

Define a *channel* $\mathcal{C}$ to be a distribution of bit sequences time stamped with monotonically non-decreasing values. The conditional distribution $\mathcal{C}_h$ describes the channel distribution conditioned on channel history h

[1] The authors of [5] are content with a stego-system with reliability $2/3$, i.e., one in which each individual bit can be incorrectly decoded with probability $1/3$, and thus require only weak error-correcting codes. However, it is clear that for a stegosystem to be useful, one would require much higher reliability. Therefore, in order to make accurate performance comparisons, we will require all stegosystems to be reliable with probability close to 1.

of previously drawn bits. All messages are assumed to be of fixed length B bits. Furthermore, assume there exists an oracle M which on input h efficiently samples the distribution $\mathcal{C}_h^B$. That is M samples $\mathcal{C}_h$ in B-bit blocks with the first bit of the block dependent on the history h and each successive bit in the block dependent on the concatenation of h and all previous bits in the block. Where the specific history h is irrelevant we will use M for $M(h)$. We also find it convenient to abbreviate the covertext distribution $\mathcal{C}_h^B$ by D except for situations where the original provides additional clarity. Additionally, we will abuse notation and use $D = \mathcal{C}_h^B$ in place of M, particularly when denoting oracle access to D and the availability or action of the sampler is not of primary interest[2].

Definition 1. A *stegosystem* or *steganographic protocol* is a pair of probabilistic polynomial time algorithms $S = (SE, SD)$ such that, for a security parameter κ,

1. SE takes as input a randomly chosen key $K \in \{0,1\}^\kappa$, a string $m \in \{0,1\}^*$ (called the ***hiddentext***), a message history h, and a channel sampling oracle $M(h)$
2. $SE^M(K, m, h)$ returns a sequence of blocks $c_1 \| c_2 \| \ldots \| c_l$ (called the ***stegotext***) each of which is an element of the support of the channel being sampled by the oracle M
3. SD takes as input a key K, a sequence of blocks $c_1 \| c_2 \| \ldots \| c_l$, a message history h, and a channel sampling oracle $M(h)$
4. $SD^M(K, c, h)$ returns a hiddentext m
5. SE^M and SD^M satisfy

$$\forall m : \Pr[SD^M(K, SE^M(K, m, h), h) = m] \geq \frac{2}{3}$$

where the randomization is over any coin tosses of SE^M, SD^M, and M.

Stegosystem Reliability. We find it useful for later discussions to elaborate the final point in the definition of a stegosystem S. The *reliability* of a stegosystem S with security parameter κ for a channel C (for messages of fixed length B) is defined as

$$\mathbf{Rel}_{\mathcal{S}(\kappa),\mathcal{C}} = \min_{m \in \{0,1\}^B} \{ \Pr_{K \in \{0,1\}^\kappa} [SD^M(K, SE^M(K, m, h), h) = m] \} .$$

While the definition of [6] considers reliability of $2/3$ sufficient, we will also consider more reliable stegosystems.

The Adversary. In this work only passive (i.e., non-corrupting) adversaries mounting chosen hiddentext attacks on S are considered. The capabilities of such an adversary are described next.

Definition 2. A probabilistic polynomial time adversary W is a (t, q, l) *passive adversary for stegosystem S with key K*, if

1. W runs in time t
2. W can make an arbitrary number of draws from C_h^B via the sampling oracle $M(\cdot)$

[2]Technically M is made necessary by the fact that we do not assume either complete knowledge of or direct access to $\mathcal{C}_h^B$. Therefore, even though we use D in place of M, throughout, we assume parties accessing D know at most the min-entropy of D.

3. W can make at most q queries of total length l bits to an oracle which is either $SE(K,\cdot,\cdot)$ or $O(\cdot,\cdot)$, where O is such that $O(m,h) \leftarrow \mathcal{C}_h^{|SE(K,m,h)|}$
4. W outputs a bit indicating whether it was interacting with SE or with O

W, sometimes referred to as the "warden"[3], is nothing more than a distinguisher in the usual sense. It attempts to differentiate between channel messages which have been altered by SE and those which have simply been drawn from C_h^B as usual and sent without modification.

Stegosystem Advantage and Insecurity. The *advantage* of a passive adversary W against stegosystem S with security parameter κ for a given channel C (for messages of fixed length B) is defined as

$$\mathbf{Adv}^{\mathrm{SS}}_{\mathcal{S}(\kappa),\mathcal{C}}(W) = \left| \Pr_{K\leftarrow\{0,1\}^\kappa;\, r\leftarrow\{0,1\}^*}[W_r^{M,SE(K,\cdot,\cdot)} = 1] - \Pr_{r\leftarrow\{0,1\}^*}[W_r^{M,O(\cdot,\cdot)} = 1] \right|$$

where the SS superscript means "Steganographic Secrecy".

For t, q, l given, the *insecurity* of stegosystem S with respect to channel C is defined as

$$\mathbf{InSec}^{\mathrm{SS}}_{\mathcal{S}(\kappa),\mathcal{C}}(t,q,l) = \max_{W\in\mathcal{W}(t,q,l)} \{\mathbf{Adv}^{\mathrm{SS}}_{\mathcal{S}(\kappa),\mathcal{C}}(W)\},$$

where $\mathcal{W}(t,q,l)$ denotes the set of all adversaries which make at most q queries of total length at most l bits and run in time at most t.

Definition 3 (Steganographic Secrecy). A stegosystem $S = (SE, SD)$ is (t,q,l,ϵ) *steganographically secret against chosen hiddentext attacks on channel distribution* $\mathcal{C}$, (t,q,l,ϵ)-SS-CHA-$\mathcal{C}$, if $\mathbf{InSec}^{\mathrm{SS}}_{\mathcal{S}(\kappa),\mathcal{C}}(t,q,l) \leq \epsilon$.

Definition 4 (Universal Steganographic Secrecy). A stegosystem S as defined above is (t,q,l,ϵ) *universally steganographically secret against chosen hiddentext attacks*, (t,q,l,ϵ)-USS-CHA, if it is (t,q,l,ϵ)-SS-CHA-$\mathcal{C}$ for all $\mathcal{C}$ satisfying $H(\mathcal{C}_h^B) > 1$ for all h drawn from $\mathcal{C}$. A stegosystem S is *universally steganographically secret* if for every channel distribution $\mathcal{C}$ and for every PPTM W, $\mathbf{Adv}^{\mathrm{SS}}_{S(\kappa),\mathcal{C}}(W)$ is negligible in κ.

With respect to the specific constructions discussed herein we need some additional notation which is also mirrors that in [6]. Let $U(k)$ denote the uniform distribution on the set of k-bit strings, and $U(B,1)$ denote the uniform distribution on predicates on B-bit strings. Let F_K, for $K \in \{0,1\}^\kappa$, denote a specific member of the family of pseudorandom predicates $\mathcal{F} : \{0,1\}^\kappa \times \{0,1\}^L \rightarrow \{0,1\}$ with key K (pseudorandom predicates and functions were first defined by [3]).

PRF Advantage and Insecurity. For a probabilistic adversary A, the *PRF-advantage of A over $\mathcal{F}$* is defined as

$$\mathbf{Adv}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(A) = \left| \Pr_{K\leftarrow U(\kappa),r\leftarrow\{0,1\}^*}[A_r^{F_K(\cdot)} = 1] - \Pr_{g\leftarrow U(L),r\leftarrow\{0,1\}^*}[A_r^g = 1] \right|.$$

For t, q given, the *insecurity* of the pseudorandom function family $\mathcal{F}$ is defined as

$$\mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(t,q) = \max_{A\in\mathcal{A}(t,q)} \mathbf{Adv}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(A),$$

where $\mathcal{A}(t,q)$ denotes the set of all adversaries which make at most q queries run in time at most t.

[3]The idea of the adversary as a warden and the use of W to designate it is a consequence of original problem formulation in [8].

Minimum Entropy. Lastly, define $H(D)$, the *minimum entropy* of probability distribution D, as

$$H(D) = \max_{x \in D} \left\{ -\log_2 \Pr_D[x] \right\}.$$

2.2 Flawed Construction 1

We now give the flawed version of Construction 1 from [6] to concretely ground later discussions of its problem and to make this work more self contained. Also, as our analysis will demonstrate, in many cases Construction 1 can be used without *any* modification, and even when modification is necessary, it will be very minimal. Construction 1, henceforth referred to as $\mathrm{S1}_{\mathrm{original}}$ for brevity, appears in Section 3.2 of [6]. $\mathrm{S1}_{\mathrm{original}}$ relies on a *rejection sampler* sub-procedure RS, which appears in Section 2.3 of [6].

This rejection sampler, and consequently $\mathrm{S1}_{\mathrm{original}}$, requires sampling access to the covertext distribution D. We denote this by oracle access to M, a sampling oracle for D ($\mathcal{C}_h^B$) or, when abusing notation, by oracle access to D itself. RS also requires access to a predicate F whose domain is the support of D. More formally, $F : \{0,1\}^B \to \{0,1\}$. RS on input a target bit y and maximum number of allowed iterations $count$ draws messages x from D until one is found that evaluates to the target bit y under F or the maximum number of allowed iterations is reached. In the latter case it outputs the last message drawn. $count$ can and should be thought of as RS's security parameter. It will become clear later that $count$ directly influences the reliability of $\mathrm{S1}_{\mathrm{original}}$. The specification of RS follows.

Procedure $\mathrm{RS}^{M,F}(y, count)$:
  $i = 0$
  repeat:
    $x \leftarrow M; i \leftarrow i + 1$
  until $F(x) = y$ or $count = i$
Output: x

The stego encoding algorithm SE for $\mathrm{S1}_{\mathrm{original}}$ takes a key K for a pseudorandom function F, additional security parameter k, hiddentext bit m, and channel history h as input[4]. It runs RS with input m and $|K|$ and outputs RS's output.

Procedure $\mathrm{S1}_{\mathrm{original}}.SE(K, k, m, h)$:
    $x \leftarrow \mathrm{RS}^{M(h),F(K,\cdot)}(m, k)$
    $h \leftarrow h \| x$
Output: x

The stego decoding or extraction algorithm SD for $\mathrm{S1}_{\mathrm{original}}$ takes the key K and a stegotext x and outputs the image of x under F as the hiddentext m.

Procedure $\mathrm{S1}_{\mathrm{original}}.SD(K, x)$:
  $m \leftarrow F(K, x)$
Output: m

From here on the sampling oracle M and message history h will cease to be explicitly mentioned when discussing RS, SE, and SD.

[4] In [6], $k = |K| = \kappa$ is a single security parameter. Here we separate them for precision of discourse.

2.3 How and Why $\mathrm{S1}_{\mathrm{original}}$ Fails

Corollary 1 in [6] falsely states that $\mathrm{S1}_{\mathrm{original}}$ is steganographically secure on all channels $\mathcal{C}$ with minimum entropy $H(D = \mathcal{C}_h^B) > 2$ against wardens W that ask only a single 1-bit query. The corollary is false as a consequence of a subtle but serious flaw in the proof of Theorem 1 which incorrectly bounds the insecurity of $\mathrm{S1}_{\mathrm{original}}$ by the insecurity of the pseudorandom function family F. The authors became aware of this issue and published [5] which contains an acknowledgment of this flaw and $\mathrm{S1}_{\mathrm{corrected}}$, a corrected version of $\mathrm{S1}_{\mathrm{original}}$, which will be discussed later.

The flaw in the proof of their Theorem 1 follows from the false implicit claim that the output of the rejection sampler using a randomly chosen predicate is identical to the covertext distribution $D = \mathcal{C}_h^B$, the input distribution for RS. This is stated more precisely and discussed in greater detail below.

***False* Claim 1.** *For any covertext distribution D with minimum entropy $H(D) > 2$, fixed bit b, randomly chosen predicate g from $U(B, 1)$, and $k \in \mathbb{N}$, the distribution of messages $x \in D$ output by $\mathrm{RS}^{D,g}(b, k)$ is identical to the distribution of messages drawn from D directly (where the probabilities are taken over the random choice of g).*

The flawed proof of the theorem tries to show, using a very straight forward two step reduction, that stegosystem $\mathrm{S1}_{\mathrm{original}}$ adversary W has advantage equal to an adversary A's advantage against the pseudorandom function F_K. In the first step, the proof shows $\mathrm{RS}^{D,F_K} \approx \mathrm{RS}^{D,g}$, and then in the second step infers $\mathrm{RS}^{D,g} = D$ using false Claim 1, and thus concludes the advantages are equal from their respective definitions. The theorem then follows directly from the respective insecurity definitions.

At first glance, false Claim 1, and consequently the flawed proof of Theorem 1, seem quite reasonable. Indeed, as the authors state, for a given bit b and randomly chosen g, it follows from the independence of D and g that $\Pr_D[x|g(x) = b : g \leftarrow U(B, 1)] = \Pr_D[x]$. However, since $\mathrm{RS}^{D,g}$ repeatedly draws blocks from D and returns the first to satisfy $g(x) = b$ *without choosing a new g* before each draw, the independence breaks down.

Nonetheless we will show that the flaw is not as bad as it may seem.

3 Bounding the Flaw

Despite the seemingly bad news that the rejections sampler perceptibly alters non-uniform covertext source distributions D, we bound the magnitude of the distortion by giving an upper bound on the statistical difference between D and $\mathrm{RS}^{D,g}$.

Before presenting the formal theorem statement, we introduce some additional notation. For a function $g : D \to \{0, 1\}$, define α_g to be the weight of g where

$$\alpha_g = \sum_{x' \in D : g(x') = 1} \Pr_D[x'] \,,$$

and β_g the weight of the complement as $\beta_g = 1 - \alpha_g$. Similarly, for a subset $S \subseteq D$, define $\alpha_S = \sum_{x' \in S} \Pr_D[x']$ and $\beta_S = 1 - \alpha_S$. Lastly, define

$$\eta(D, k) = \frac{1}{2^{|D|}} \sum_{S \subsetneq D} \alpha_S^k$$

and

$$\zeta(D, k) = \frac{1}{2^{|D|}} \sum_{S \subseteq D} \alpha_S^k = \eta(D, k) + \frac{1}{2^{|D|}} \,,$$

Note that, for a fixed D, $\eta(D,k)$ is a negligible function of k (provided D has no zero-probability elements), because $\alpha_S < 1$ for $S \subsetneq D$.

Theorem 1. *Let D be any discrete probability distribution, $k \in \mathbb{N}$ and a bit $b \in \{0,1\}$. Let p be the probability of the most likely event in D. Then for a randomly chosen predicate $g : D \to \{0,1\}$, the statistical difference between D and $\mathrm{RS}^{D,g}(b,k)$ is at most $2p$ plus a negligible function in k. More precisely,*

$$\sum_{\forall x \in D} \left| \Pr_D[x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] \right| \leq 2p + 2\eta(D,k)\,.$$

The remainder of this section is devoted to formulating and proving a number of intermediate results which when taken together will yield the proof of Theorem 1.

3.1 Supporting Results

On the way to proving Theorem 1, the first step is to quantify the output distribution of the rejection sampler. First we consider the limiting case when the maximum number of allowed channel draws made by RS, the parameter k in the above, is allowed to go to infinity. Note that in $\mathrm{S1}_{\mathrm{original}}$, the security parameter k, which is length of the pseudorandom function key K, is also used as the cutoff parameter for RS. However, from here on k will only denote the maximum number of attempts made by $RS[]$, and κ will denote the security parameter for $\mathrm{S1}_{\mathrm{original}}$ and the length of the pseudorandom function key K. The following lemma provides an expression for the probability distribution of RS in the infinite case. Lemma 2 then uses this expression to give a version of Theorem 1 in the case of an infinite k.

Lemma 1. *For x an element from the support of D and a bit $b \in \{0,1\}$, let us define $\mathrm{RS}^{D,g}(b,\infty) \equiv \lim_{k\to\infty} \mathrm{RS}^{D,g}(b,k)$ and $\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] \equiv \lim_{k\to\infty} \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x]$. Then,*

$$\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] = \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{g\in U(B,1):g(x)=1} \frac{1}{\alpha_g}\right)$$

where the probability is taken over the choice of g.

Proof. We will prove the case of $b = 1$ and argue by symmetry that this also suffices to prove the case of $b = 0$. To compute the probability that $\mathrm{RS}^{D,g}(1,k)$ outputs x, simply find the expected value over the $2^{|D|}$ possible random functions $g : D \to \{0,1\}$, as follows,

$$\begin{aligned}
\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1,k) \to x] &= \frac{1}{2^{|D|}} \left(\sum_{g:g(x)=1} \Pr_D[x] \sum_{i=0}^{k-1} \beta_g^i + \sum_{g:g(x)=0} \Pr_D[x] \beta_g^{k-1} \right) \\
&= \frac{\Pr_D[x]}{2^{|D|}} \left(\sum_{g:g(x)=1} \frac{1-\beta_g^k}{1-\beta_g} + \sum_{g:g(x)=0} \beta_g^{k-1} \right). \qquad (1)
\end{aligned}$$

Taking the limit as $k \to \infty$, that is as the rejection sampler makes greater and greater numbers of draws from D before "giving up", we have

$$\begin{aligned}
\lim_{k\to\infty} \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1,k) \to x] &= \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{g:g(x)=1} \frac{1}{1-\beta_g}\right) \\
&= \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{g:g(x)=1} \frac{1}{\alpha_g}\right).
\end{aligned}$$

It remains to prove the case for $b = 0$. However, by symmetry, for each specific function g which maps an element x to 0, there exists a unique $\hat{g}$ such that $\forall x \in D, \hat{g}(x) = 1 - g(x)$. Consequently, for each function g we have,

$$\Pr[\mathrm{RS}^{D,g}(0,k) \to x] = \Pr[\mathrm{RS}^{D,\hat{g}}(1,k) \to x].$$

Generalizing this over all possible choices for the function g gives

$$\Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(0,k) \to x] = \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(1,k) \to x]$$

so our consideration of $\mathrm{RS}^{D,g}(1,k)$ is sufficient and the proof is complete. □

Remark 1. It can be seen from (2) and some algebra, that when $k = 2$, in fact, $\Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] = \Pr_D[x]$ as stated in [5]. Indeed, the proposed fix in [5] is to set $k = 2$ and accept the fact that this causes a high probability (between $1/4$ and $3/8$) of decoding incorrectly, and thereby reduced reliability.

Now we give the infinite analog of Theorem 1 which we use later in its proof.

Lemma 2. *Let D be any discrete probability distribution and $b \in \{0,1\}$ a bit. Let p be the probability of the most likely event in D. Then for a randomly chosen predicate $g : D \to \{0,1\}$, the statistical difference between D and $\mathrm{RS}^{D,g}(b,\infty)$ is at most $2p$. More precisely,*

$$\sum_{\forall x \in D} \left| \Pr_D[x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] \right| \leq 2p\,.$$

The proof employs the following proposition which is a consequence of the relationship between the harmonic and arithmetic means.

Proposition 1. *For a set of n non-zero real numbers $a_1, a_2, \ldots, a_n$,*

$$\frac{1}{a_1} + \cdots + \frac{1}{a_n} \geq \frac{n^2}{(a_1 + \cdots + a_n)}\,.$$

Proof. The proposition can be verified by recalling that the *harmonic mean* of a set of n values $a_1, a_2, \ldots, a_n$, is defined as $n/(1/a_1 + \cdots + 1/a_n)$, whereas the usual *arithmetic mean* is defined as $(a_1 + \cdots + a_n)/n$. A well known property of the harmonic mean is that it is less than or equal to the arithmetic mean for the same set of numbers with equality only when all a_i are equal [1]. Therefore, inverting both sides of this relation and multiplying by n, gives the above proposition. □

Proof of Lemma 2. First we remind the reader of the property of the statistical difference that for any distributions D_1 and D_2,

$$\sum_{\forall x \in D_1, D_2} \left| \Pr_{D_1}[x] - \Pr_{D_2}[x] \right| = 2 \sum_{x \in D_1, D_2 : \Pr_{D_1}[x] \geq \Pr_{D_2}[x]} \Pr_{D_1}[x] - \Pr_{D_2}[x]\,.$$

For the remainder of the proof, where not indicated probabilities are with respect to D. Also, define $t = |D|$.

For each function g, let us consider the subset S of D which is the pre-image of 1 under g, that is $S = \{x \in D : g(x) = 1\}$. Since there are 2^{t-1} subsets S containing any given element x, rewriting

Lemma 1 in terms of S rather than g and applying the inequality of Proposition 1 to the result gives,

$$\begin{aligned}
\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x] &= \frac{\Pr[x]}{2^t}\left(1+\sum_{S\subseteq D:x\in S}\frac{1}{\alpha_S}\right)\\
&\geq \frac{2^{2(t-1)}\Pr[x]}{2^t\sum\limits_{S\subseteq D:x\in S}\alpha_S}\\
&= \frac{2^{t-2}\Pr[x]}{\sum\limits_{S\subseteq D:x\in S}\sum\limits_{\forall x\in S}\Pr[x]}\\
&= \frac{2^{t-2}\Pr[x]}{2^{t-1}\Pr[x]+2^{t-2}\sum\limits_{x'\neq x}\Pr[x']}\\
&= \frac{\Pr[x]}{2\Pr[x]+1-\Pr[x]} = \frac{\Pr[x]}{1+\Pr[x]}\,.
\end{aligned}$$

Thus,

$$\begin{aligned}
\Pr_D[x]-\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x] &\leq \Pr[x]-\frac{\Pr[x]}{1+\Pr[x]}\\
&= \frac{(\Pr[x])^2}{1+\Pr[x]}\\
&\leq (\Pr[x])^2\,.
\end{aligned}$$

Finally, combining these two pieces,

$$\begin{aligned}
&\sum_x\left|\Pr_D[x]-\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x]\right|\\
&= 2\sum_{\{x:\Pr[x]\geq\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x]\}}\Pr[x]-\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x]\\
&\leq 2\sum_{\{x:\Pr[x]\geq\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x]\}}(\Pr[x])^2\\
&\leq 2\sum_{\forall x\in D}(\Pr[x])^2\\
&\leq 2p\sum_{\forall x\in D}\Pr[x]=2p\,,
\end{aligned}$$

where p is the probability of the most probable element in D. □

Lastly, we consider the statistical difference between the probability distributions of the finite and infinite rejection samplers.

Lemma 3. *For a fixed $k\in\mathbb{N}$,*

$$\sum_{\forall x\in D}\left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty)\to x]-\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,k)\to x]\right|\leq 2\eta(D,k)$$

Proof. Using (2) from the proof of Lemma 1 it follows that

$$\sum_{\forall x \in D} \left| \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] \right| \tag{2}$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| 1 + \sum_{S \subseteq D : x \in S} \frac{1}{\alpha_S} - \sum_{S \subseteq D : x \in S} \frac{\beta_S^k - 1}{\beta_S - 1} - \sum_{S \subseteq D : x \notin S} \beta_S^{k-1} \right| \tag{3}$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| 1 + \sum_{S \subseteq D : x \in S} \frac{1}{\alpha_S} - \sum_{S \subseteq D : x \in S} \frac{1 - \beta_S^k}{\alpha_S} - \sum_{S \subseteq D : x \in S} \alpha_S^{k-1} \right| \tag{4}$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| \sum_{S \subsetneq D : x \in S} \frac{\beta_S^k - \alpha_S^k}{\alpha_S} \right| \tag{5}$$

$$= 2 \sum_{x \in D : |\cdot| \geq 0} \frac{\Pr[x]}{2^{|D|}} \sum_{S \subsetneq D : x \in S} \frac{\beta_S^k - \alpha_S^k}{\alpha_S} \tag{6}$$

$$\leq \frac{1}{2^{|D|-1}} \sum_{\forall x \in D} \Pr[x] \sum_{S \subsetneq D : x \in S} \frac{\beta_S^k}{\alpha_S} \tag{7}$$

$$= \frac{1}{2^{|D|-1}} \sum_{S \subsetneq D : S \neq \emptyset} \frac{\beta_S^k}{\alpha_S} \sum_{\forall x \in S} \Pr[x] \tag{8}$$

$$= \frac{1}{2^{|D|-1}} \sum_{S \neq \emptyset} \beta_S^k = \frac{1}{2^{|D|-1}} \sum_{S \subsetneq D} \alpha_S^k \tag{9}$$

$$\leq 2\eta(D,k)\,. \tag{10}$$

Line (4) follows from the definitions of α and β and the symmetry of the set of all functions. To obtain Line (5), combine the sums and remove the term 1 by restricting S to be a *proper* subset of D. Line (6) follows from the same property of statistical difference used in the proof of Lemma 2. Line (8) follows by expanding the sums, gathering common terms with respect to a specific subset S and rewriting the sums with the appropriate modifications to their bounds (the empty set is excluded because every subset S must have at least one element). Canceling the α_S denominator and noting that $\beta_D = \alpha_\emptyset = 0$ gives us the last line and completes the proof. □

3.2 Putting It Together

At this point we have assembled the necessary tools to prove our bound on the statistical difference between an arbitrary message distribution D and $\mathrm{RS}^{D,g}(b,k)$ for a random function g.

Proof of Theorem 1. The proof follows by first inserting positive and negative $\Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x]$ inside the absolute value signs, applying the triangle inequality, and then using Lemmas 2 and 3 . That is,

$$\sum_{\forall x \in D} \left| \Pr_D[x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] \right|$$

$$= \sum_{\forall x \in D} \left| \Pr_D[x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] \right.$$

$$
\begin{aligned}
& + \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] \Bigg| \\
\leq & \sum_{\forall x \in D} \left| \Pr_{D}[x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] \right| \\
& + \left| \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x] \right| \\
\leq & \; 2p + 2\eta(D,k)\,.
\end{aligned}
$$

□

4 Fixing the Flaw

Recall that the minimum entropy $H(D)$ is $-\log_2 p$, where p is the highest probability of an event in D. Thus, the bound of Theorem 1 shows that the flaw (i.e., the statistical difference between the output of the rejection sampler and D) is exponentially small in $H(D)$, plus a negligible amount: $2p + 2\eta(D,k) = 2^{1-H(D)} + 2\eta(D,k)$. Therefore, we have shown that the following is true.

Observation 1. For D with sufficiently high min-entropy, $\mathrm{S1}_{\mathrm{original}}$ (i.e., Construction 1 of [6]) needs no modification.

On the other hand, since p is fixed for any given D, when D *lacks* sufficiently high min-entropy, $\mathrm{S1}_{\mathrm{original}}$ in its current form is *not* steganographically secret, that is it is insecure. This brings us to our second main contribution: a modified version of $\mathrm{S1}_{\mathrm{original}}$ that is secure *for all* D. We call it MESS for "Minimum-Entropy-Sensitive Stegosystem."

4.1 Our Construction

The problem with $\mathrm{S1}_{\mathrm{original}}$ is that it is stuck with whatever min-entropy D provides. To fix this, we propose RS-HE, a modified version of RS, that uses repeated sampling on D to effectively increases the minimum entropy. Specifically, instead of using one covertext message $x \in D$ per hiddentext bit, RS-HE uses a variable number of covertexts $x_i \in D$. The concatenation of all of these x_i is then evaluated under the predicate F (with a suitably expanded domain). The exact number of covertexts which make up the stegotext depends directly on $H(D)$ and is fixed for a given D. Our proposed stegosystem MESS is the same as $\mathrm{S1}_{\mathrm{original}}$ except for a few minor syntactic changes necessary to accommodate its use of RS-HE instead of RS.

4.1.1 The Memoryless Channel Case

For now, assume that the channel is memoryless: D is independent of the previous message history h. In other words, successive covertext messages are independent of one another. Consequently h can be completely ignored and is suppressed.

Let n be an additional security parameter for MESS and RS-HE. It specifies the number elements of D (covertexts) over which a single hiddentext bit will be encoded. Recall that $\mathrm{S1}_{\mathrm{original}}$ and RS had security parameters $\kappa = |K|$ and k, the length of the pseudorandom predicate key and the number of attempts made by RS respectively. As before, in general, RS-HE uses a predicate F, but the domain is expanded, i.e. now $F : D^n \to \{0,1\}$. When running as a subroutine of MESS, RS-HE has oracle access to F_K, a specific pseudorandom predicate family member with key $K \in \{0,1\}^\kappa$.

The modified version of RS-HE is:

Procedure RS-HE$^{D,F}(y, count, n)$:
$i = 0$
repeat:
for $j = 1$ to n:
$x_j \leftarrow D$
$x \leftarrow (x_1 \parallel x_2 \parallel \ldots \parallel x_n)$
$i \leftarrow i + 1$
until $F_K(x) = y$ or $count = i$
Output: x

The only differences between the stego-encoding algorithms of MESS and $\mathrm{S1}_{\mathrm{original}}$ is that MESS.SE has additional input n that it uses when it calls RS-HE, and its stegotext output is n times longer. The stego-decoding algorithm MESS.SD is unchanged from $\mathrm{S1}_{\mathrm{original}}.SD$ except that its stegotext input is n times longer. It should be emphasized that with respect to the "flawed" $\mathrm{S1}_{\mathrm{original}}$ given in Section 2.2, the only differences in MESS (aside from those between RS-HE and RS) are the additional security parameter n input to both SE and SD, the expansion of the domain of F_K, and the n times longer stegotext output by SE and input to SD.

4.1.2 The General Case

To generalize our modifications, we drop the memoryless channel assumption. Suppose instead that the distribution of covertexts *does* depend on the history h of previously sent messages. In other words, D truly is conditioned by h. The distribution resulting from sending n messages is more complex than D^n. Let $D^{(n)}$ denote this distribution. With respect to the original channel notation, $D^{(n)} \equiv \mathcal{C}_h^{nB}$ (recall that $\mathcal{C}_h^{nB}$ denotes a conditional distribution of messages of fixed length nB bits conditioned on history h). The general version of RS-HE then is:

Procedure RS-HE$^{M,F}(y, count, n)$:
$i = 0$
repeat:
for $j = 1$ to n:
$x_j \leftarrow M(h)$
$h \leftarrow h \parallel x_j$
$x \leftarrow (x_1 \parallel x_2 \parallel \ldots \parallel x_n)$
$i \leftarrow i + 1$
until $F_K(x) = y$ or $count = i$
Output: x

The resulting MESS.SE and MESS.SD are the same as described for D^n in Section 4.1.1 with the stipulation that now $F_K : D^{(n)} \rightarrow \{0, 1\}$.

Remark 2. The inner "for" loop of RS-HE can be thought of as an oracle $M^{(n)}$—an efficient sampling oracle for $D^{(n)}$. Observe that such a sampling oracle can be always be built given n and access to the original oracle M. Thus, the analysis of RS given in Theorem 1 applies here as well, except that D must be replaced with $D^{(n)}$.

4.2 Proof of Correctness

The proof of $\mathrm{S1}_{\mathrm{original}}$ given in [6] only attempted to show security with respect to adversaries making a single 1-bit query; we will do the same. The techniques of [6] for going from 1-bit to multi-bit stegosystems (namely, maintaining state in the form of a counter) apply here as well. It remains an interesting problem to construct better stegosystems for multi-bit messages.

The proof that MESS is 1-bit steganographically secure follows from Theorem 1 with $D^{(n)}$ in place of D. Clearly the first term becomes at most p^n and can be made negligible by taking n sufficiently large. The only complication is that the second term, $\eta(D^{(n)}, k) = 2^{-|D^{(n)}|} \sum_{S \subsetneq D^{(n)}} \alpha_S^k$ now depends on both n and k. We need to show that it can be made negligible even as n grows. We do this after stating and prior to proving the security theorem for MESS.

Let $\mathrm{MESS}(\kappa, k, n)$ denote our new system instantiated with security parameters κ (key length for PRF F), k (number of tries before RS-HE gives up) and n (hiddentext stretch factor).

Theorem 2. *Let D be a covertext message distribution conditioned on message history h, and let p be the probability of the most likely element of D ($p = 2^{-H(D)}$). Then*

$$\mathbf{InSec}^{\mathrm{SS}}_{\mathit{MESS}(\kappa,k,n),D}(t,1,1) \le 2\left(p^n + \left(\frac{3}{4}\right)^k + \mathrm{e}^{-\lfloor \frac{1}{p^n} \rfloor \frac{1}{8}}\right) + \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(t + O(k), k).$$

More generally, for any $0 < \delta < 1/2$,

$$\mathbf{InSec}^{\mathrm{SS}}_{\mathit{MESS}(\kappa,k,n),D}(t,1,1) \le 2\left(p^n + \left(\frac{1}{2} + \delta\right)^k + \mathrm{e}^{-\lfloor \frac{1}{p^n} \rfloor 2\delta^2}\right) + \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(t + O(k), k)$$

(the first formula is simply an instantiation of the second with $\delta = 1/4$).

Before proving Theorem 2 we deal with the issue of bounding $\eta(D^{(n)}, k)$ in two steps. It is easier to bound a closely related value

$$\zeta(D^{(n)}, k) = \frac{1}{2^{|D^{(n)}|}} \sum_{S \subseteq D^{(n)}} \alpha_S^k = \eta(D^{(n)}, k) + \frac{1}{2^{|D^{(n)}|}},$$

which differs from η only by the inclusion of the full subset $S = D^{(n)}$ in the sum. As we will see in Section 5, ζ is exactly the failure probability of the rejection sampler.

Lemma 4 bounds $\zeta(D, k)$, for any distribution D, by $\zeta(U_D, k)$, where U_D is the uniform distribution with essentially the same min-entropy as D. Lemma 5 bounds ζ of this uniform distribution.

Lemma 4. *Among all distributions of a given min-entropy, ζ is the largest for the uniform distribution. More precisely, for a distribution D with minimum entropy $H(D)$, define $U_D = U(\lfloor 2^{H(D)} \rfloor)$, that is U_D is a uniform distribution with $\lfloor 2^{H(D)} \rfloor$ elements. Then for all $k \in \mathbb{N}$, $\zeta(D, k) \le \zeta(U_D, k)$*

The following two claims will help with the proof of Lemma 4.

Claim 1. *If D has an element with zero probability and D' differs from D only by the removal of this zero probability element, then $\zeta(D', k) = \zeta(D, k)$.*

Proof. This is easily verified using the definition of ζ: the number of terms in the sum is cut in half (with every pair of terms of equal weight becoming one), but the coefficient in front of the sum is multiplied by two. □

Claim 2. *Let a, b be elements of D with probabilities p_a and p_b such that $p_a \geq p_b$. Define D'' to be the distribution with the same probabilities as D except with $p_a+\gamma$ and $p_b-\gamma$ in place of p_a and p_b respectively ($0 \leq \gamma \leq p_b$). Then $\zeta(D'', k) \geq \zeta(D, k)$.*

Proof. For $\gamma = p_b$, a simple proof is obtained by using the definition of ζ to rewrite the two expressions as sums. Then using binomial series and regrouping the terms the claim follows directly. For the general case one can treat $\zeta(D'', k)$ as a continuous real-valued function of γ. Then

$$\zeta(D''(\gamma), k) = \frac{1}{2^{|D|}} \sum_{S \subset D: a, b \notin S} (\alpha_S + p_a + \gamma)^k + (\alpha_S + p_b - \gamma)^k + \alpha_S^k + (\alpha_S + p_a + p_b)^k \,.$$

Taking the derivative with respect to γ we obtain

$$\frac{k}{2^{|D|}} \sum_{S \subset D: a, b \notin S} (\alpha_S + p_a + \gamma)^{k-1} - (\alpha_S + p_b - \gamma)^{k-1} > 0 \,,$$

because $p_a > p_b \geq \gamma$. Hence $\zeta(D'', k)$ is a nondecreasing function of γ on the interval $0 \leq \gamma \leq p_b$. □

Proof of Lemma 4. We can transform D into U_D by adding the mass to the highest-probability elements until their probability reaches $1/\lfloor 2^{H(D)} \rfloor$, while simultaneously removing the same mass from lowest-probability elements until their probability reaches 0. By Claim 2, ζ of the resulting distribution will not decrease. Then we remove all zero-probability elements to obtain U_D (this, by Claim 1, will not change ζ). □

Lemma 5. *For $U(t)$, a uniform distribution on t elements, $\zeta(U(t), k)$ can be made negligible for both t and k sufficiently large. Specifically for $0 < \delta < \frac{1}{2}$, $\zeta(U(t), k) \leq \left(\frac{1}{2} + \delta\right)^k + \mathrm{e}^{-2t\delta^2}$.*

Proof. Consider ζ as a subset of a union of two "bad" events: (1) that fewer than $1/2 + \delta$ elements of $U(t)$ map to 1 under g or (2) that more than $1/2 + \delta$ elements of $U(t)$ map to 1 under g, but not one of those gets selected after k tries. More precisely, rewriting the definition of ζ,

$$\begin{aligned}
\zeta(U(t), k) &= \sum_{\forall S \subseteq U(t)} \frac{\alpha_S^k}{2^{|t|}} \\
&= \left[\Pr[\alpha_S \leq (1/2 + \delta)] \sum_{S: \alpha_S \leq (1/2+\delta)} \alpha_S^k \right] + \left[\Pr[\alpha_S > (1/2 + \delta)] \sum_{S: \alpha_S > (1/2+\delta)} \alpha_S^k \right] \\
&\leq \left(\frac{1}{2} + \delta \right)^k + \mathrm{e}^{-2t\delta^2} \,.
\end{aligned}$$

The exponential term follows from the application of Hoeffding's Inequality[5] [4] to $\Pr_g[\alpha_S > (1/2+\delta)] = \Pr_g[t\alpha_S > t(1/2+\delta)]$. It is a Chernoff like bound which states that for t independent $0/1$ random variables X_i each with probability p, the random variable $S = \sum_{i=1}^t X_i$ obeys,

$$\Pr[S \geq pt + \delta t] \leq \mathrm{e}^{-2t\delta^2} \,.$$

□

[5]The use of such a bound makes sense since for $S \subset U(t)$, $t\alpha_S = |S|$, that is the number of heads/ones observed for on t independent fair coin tosses.

Proof of Theorem 2. We first consider the case of MESS for a truly random predicate F and then add the necessary correction for a pseudorandom F. The security of MESS is completely determined by the security of RS-HE and the pseudorandom random predicate F which it accesses.

Recall that $D^{(n)}$ is the covertext distribution consisting of n subsequent draws from the given covertext distribution D via its sampling oracle $M(h)$ with message history input h. Let $M^{(n)}(h)$ be an efficient sampling oracle for $D^{(n)}$. As we pointed out in the remark at the end of Section 4.1.2, such an $M^{(n)}$ can be easily constructed from M and, in fact, RS-HE$^{M(\cdot),F}(b,k)$ is equivalent to $\mathrm{RS}^{M(\cdot),F}(b,k)$ for the same predicate F. Thus applying Theorem 1 gives,

$$\begin{aligned}
&\sum_{\forall x \in D^{(n)}} \left| \Pr_{D^{(n)}}[x] - \Pr_{F \in U(nB,1),M}[\text{RS-HE}^{M(\cdot),F}(b,k) \to x] \right| \\
&= \sum_{\forall x \in D^{(n)}} \left| \Pr_{D^{(n)}}[x] - \Pr_{F \in U(nB,1),M}[\mathrm{RS}^{M^{(n)}(\cdot),F}(b,k) \to x] \right| \\
&\leq 2p^n + 2\eta(D^{(n)},k) \qquad (11)
\end{aligned}$$

where as previously defined, p is the largest probability in D and $\eta(D^{(n)},k) = 2^{-|D^{(n)}|} \sum_{S \subsetneq D^{(n)}} \alpha_S^k$.

Clearly the first term in 11 can be made negligible since n is now a system parameter. It remains to show that even with the added dependency on n, $\eta(D^{(n)},k)$ can also be made negligible. Using Lemma 4 and Lemma 5 with $t = \lfloor p^{-n} \rfloor$ we have

$$\begin{aligned}
\eta(D^{(n)},k) &< \zeta(D^{(n)},k) \\
&\leq \left(\frac{1}{2}+\delta\right)^k + \mathrm{e}^{-\lfloor p^{-n} \rfloor 2\delta^2} \qquad (12)
\end{aligned}$$

Finally, combining (11) and (12) and accounting for the advantage due to a pseudorandom F,

$$\mathbf{Adv}^{\mathrm{SS}}_{\mathrm{MESS}_{(\kappa,k,n),D}}(W) \leq 2p^n + 2\left(\frac{1}{2}+\delta\right)^k + 2\mathrm{e}^{-\lfloor p^{-n} \rfloor 2\delta^2} + \mathbf{Adv}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(A)\,,$$

where $0 < \delta < 1/2$. Therefore by the definition of insecurity,

$$\mathbf{InSec}^{\mathrm{SS}}_{\mathrm{MESS}_{(\kappa,k,n),D}}(t,1,1) \leq 2\left(p^n + \left(\frac{1}{2}+\delta\right)^k + \mathrm{e}^{-\lfloor p^{-n} \rfloor 2\delta^2}\right) + \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(t + O(k), k)\,.$$

□

5 Performance

5.1 Reliability

We provided an explicit bound on the insecurity $\mathbf{InSec}$ of our stegosystem MESS in the previous section. However, there is another important stegosystem property: reliability $\mathbf{Rel}$, that is, the probability that the recipient decodes the encoded message correctly. While Definition 1 requires only $\mathbf{Rel} \geq 2/3$, in reality the communicating parties will most likely desire $\mathbf{Rel} \approx 1$. We bound the reliability of MESS in the following theorem.

Theorem 3. *Let D be a covertext message distribution conditioned on message history h with $H(D) > 1$ and let p be the probability of the most likely element of D ($p = 2^{-H(D)}$). Then for any $0 < \delta < \frac{1}{2}$,*

$$\mathbf{Rel}_{MESS(\kappa,k,n)} \geq 1 - \left(\left(\frac{1}{2} + \delta \right)^k + \mathrm{e}^{-\lfloor \frac{1}{p^n} \rfloor 2\delta^2} \right) - \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k) \,.$$

Lemma 6. *For any distribution D and bit $b \in \{0, 1\}$, for a randomly chosen predicate $F \leftarrow U(|D|, 1)$, the encoding error introduced by $\mathrm{RS}^{D,F}(b, k)$ is equal to $\zeta(D, k)$, where $\zeta(D, k) = \frac{1}{2^{|D|}} \sum_{S \subseteq D} \alpha_S^k$ as previously defined.*

Proof. $\mathrm{RS}^{D,F}(b, k)$ introduces encoding error whenever after k unsuccessful attempts to find a covertext $x \in D$ such that $F(x) = b$, it outputs the last (kth) x drawn from D. Using algebra similar to that in the proof of Lemma 1, this probability can be shown to be $\zeta(D, k)$. □

Proof of Theorem 3. The reliability of $\mathrm{MESS}(\kappa, k, n)$ is simply one minus the encoding error introduced by $\text{RS-HE}^{D,F_K}(\cdot, k, n)$ where $F_K \in \mathcal{F}(\kappa)$, now a pseudorandom predicate family with security parameter κ on the domain $D^{(n)}$. Recall that in the proof of Theorem 2 it was argued that $\text{RS-HE}^{D,F_K}(\cdot, k, n)$ and $\mathrm{RS}^{D^{(n)},F_K}(\cdot, k)$ are equivalent (see also the Remark of Section 4.1.2). So, by Lemma 6 and the definition of pseudorandom function insecurity, the encoding error introduced by $\text{RS-HE}^{D,F_K}(\cdot, k, n)$ is at most $\zeta(D^{(n)}, k) + \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k)$ (the $O(nk)$ is because the running time of the rejection sampler, which is playing the role of the "adversary" here, is $O(nk)$, not counting time required for answering queries to D and the PRF). Using the upper bound for $\zeta(D^{(n)}, k)$ from (12) in the proof of Theorem 2 and subtracting from one gives the indicated lower bound for the reliability. □

5.2 Parameter Choice for MESS

Given covertext distribution D with min-entropy $H(D) > 1$, for MESS to operate with 2^{-80} security and a corresponding reliability of at least $1 - 2^{-80}$, what values of the parameters κ, k, and n are necessary? We want each of the four terms in Theorem 2 to be less than 2^{-82}. Substituting $2^{-H(D)}$ for p, the most likely element in D, in the first term and solving gives

$$nH(D) \geq 83 \,. \tag{13}$$

Solving the second term for k where $0 < \delta < 1/2$ gives,

$$k \geq \frac{-83}{\log_2(1/2 + \delta)} \,. \tag{14}$$

Substituting $2^{-H(D)}$ for p in the third term and solving for $nH(D)$, again where $0 < \delta < 1/2$, gives,

$$nH(D) \quad \geq \quad \log_2 \left(\frac{83}{2\delta^2 \log_2 \mathrm{e}} + 1 \right) \,. \tag{15}$$

Since n must satisfy both (13) and (15) and we want n to be as small as possible, we take $n \geq \lceil 83/H(D) \rceil$ as specified by (13). Straightforward calculation reveals that as long as $\delta \geq 2^{-39}$ the right hand side of (15) will be at most 83. Substituting this value of $\delta = 2^{-39}$ in (14) gives the constraint $k \geq 84$. Finally, κ is chosen so that the insecurity $\mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k)$ of the given PRF family $\mathcal{F}$ is at most 2^{-82}. These same parameter choices will also provide the desired reliability level.

Final Values: For 2^{-80} security and a corresponding reliability of at least $1 - 2^{-80}$, MESS requires $n \geq \lceil 83/H(D) \rceil$, $k \geq 84$, and κ such that for the chosen PRF family $\mathcal{F}$, $\mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k) \leq 2^{-82}$.

5.3 Efficiency

When talking about the efficiency of a stegosystem S it makes sense to consider the respective running times of the stego-encoder and decoder algorithms, $S.SE$ and $S.SD$, and the length of the stegotext output by $S.SE$. These metrics are functions of both the required level of security *and* the required reliability level. Given a particular reliability level, we can consider the resulting "noisy channel" created by the stegosystem for sending hiddentext bits. An important characteristic of this channel is its rate (which will be inversely proportional to the stegotext length).

For each hiddentext bit, the stego-encoder for our MESS essentially just draws, on average, $2n$ samples from the covertext distribution D and evaluates twice the pseudorandom predicate F_K on the concatenation of n samples. Similarly, for each hiddentext bit, our stego-decoder just evaluates F_K on the stegotext received, i.e., on the concatenation of the n messages from D. So, the running times of our decoder is essentially one PRF evaluation, and the average running time of our encoder is about twice that. The stegotext length is just n covertexts long. With reliability as high as $1 - 2^{-80}$, our hiddentext transmission rate will be essentially 1 hiddentext bit per n covertexts.

Remark 3. As the analysis of the previous section indicates, for 2^{-80} security and $1 - 2^{-80}$ reliability, *our stegotext is only one covertext long* if $H(D) \geq 83$.

Remark 4. Although k is high, it does not affect performance much, as it is only the *maximum* number of attempts RS-HE makes. The *expected* number of attempts RS-HE makes is just 2.

Remark 5 (Multi-bit Efficiency:). MESS can be made to work for multi-bit hiddentext messages through the addition of a stateful counter N (both the sender and receiver must keep it synchronized) that is input to F_K along with the n concatenated covertexts (see [6] or [5] for details). The per hiddentext bit efficiency remains unchanged.

6 Performance Comparison: MESS versus $\mathrm{S1}_{\mathrm{corrected}}$

After sketching Hopper, Langford, and van Ahn's revised Construction 1 which appears in [5], hereafter denoted $\mathrm{S1}_{\mathrm{corrected}}$, we juxtapose it with our construction MESS. We compare the schemes in two cases: 1) when they are parameterized by $H(D)$ and 2) when they operate as universal stegosystems, that is without any knowledge of D (aside from that gained by sampling it). For simplicity, we always assume that $H(D) \geq 1$, i.e., no element of D has probability greater than $1/2$. Recall also that both schemes assume the covertext distribution D is over fixed-length messages, so it will be convenient to measure stegotext length and hiddentext transmission rate (per hiddentext bit) in terms of the number of covertext messages.

6.1 Revised Construction 1

Hopper, Langford, and van Ahn corrected the flaw of $\mathrm{S1}_{\mathrm{original}}$, described in Section 2.3 of this work, shortly after its publication in [6]. They gave $\mathrm{S1}_{\mathrm{corrected}}$, a revised version, in [5], but did so without any analysis of the severity of the flaw.

There are two main differences between $\mathrm{S1}_{\mathrm{corrected}}$ and $\mathrm{S1}_{\mathrm{original}}$. First, although $\mathrm{S1}_{\mathrm{corrected}}$ uses the same rejection sampler RS as $\mathrm{S1}_{\mathrm{original}}$ did, it forces RS to give up after only $k = 2$ attempts. In this case the output distribution of RS can be shown, as in [5] or using our Lemma 1, to be identical to the covertext distribution D. Unfortunately, as the authors point out, limiting RS to 2 attempts increases the probability Δ that an encoding error is introduced by $\mathrm{RS}^{D,F_K}(b, 2)$ to $\Delta = \frac{1}{2} - \frac{1-p}{4}$ (plus the PRF insecurity), where p is the highest probability in D. So, depending on the covertext distribution D, $1/4 < \Delta \leq 3/8$, where the upper bound of $3/8$ comes from the assumption that $H(D) \geq 1$. Essentially, the encoding error increases

because there is a good chance the rejection sampler will not find a covertext $x \in D$ such that $F_K(x) = b$ in just two tries. This motivates the second main difference: the use of an error-correcting code by $\mathrm{S1}_{\mathrm{corrected}}$. In order to achieve reliable (i.e. $\mathbf{Rel} \approx 1$) hiddentext transmission, prior to stego-encoding $\mathrm{S1}_{\mathrm{corrected}}$ must first encode the hiddentext input using an error correcting code that corrects Δ fraction of errors. The stego-decoder $\mathrm{S1}_{\mathrm{corrected}}.SD$, in turn, as its final step reconstructs the transmitted hiddentext from the error-encoding codewords it recovered.

We note that both $\mathrm{S1}_{\mathrm{corrected}}$ and MESS can securely and reliably send multi-bit hiddentext messages through the use of a stateful counter N (both the sender and receiver must keep it synchronized) that is an addition input to F_K (see [5] for details).

6.2 Distribution-Dependent Comparison

We consider the relative performance of $\mathrm{S1}_{\mathrm{corrected}}$ and MESS for 2^{-80} security and $1 - 2^{-80}$ reliability.

The error correcting codes needed by $\mathrm{S1}_{\mathrm{corrected}}$ to assure reliable hiddentext transmission[6] will stretch each hiddentext message bit by a code-dependent factor $\ell = 1/R$, where R is the rate of the code. Note that the "noisy channel" created by the error-prone stego-encoder is essentially a binary symmetric channel with bit-flip probability Δ, and therefore the rate R of the code is bounded by the channel capacity $C = 1 - H_2(\Delta)$, where $H_2(\Delta)$ denotes the binary entropy of the distribution $(\Delta, 1-\Delta)$. Plugging in the bounds on Δ gives us

$$\frac{1}{5} \approx 1 - H_2(1/4) > C > 1 - H_2(3/8) \approx \frac{1}{22}\,,$$

Therefore, depending on the min-entropy of D, a single hiddentext bit will be stretched to between 5 and 22 bits, *each of which* must then be stego-encoded using RS^{D,F_K}. Thus for secure and reliable hiddentext transmission by $\mathrm{S1}_{\mathrm{corrected}}$ the total stegotext will be between 5 and 22 covertext messages long per hiddentext bit. (Note that this holds asymptotically, i.e., when multiple hiddentext bits are sent and optimal codes are used; for sending just one or a few bits, the situation is even worse, because more expensive codes must be used to achieve reliability for short hiddentext messages). This means that the effective hiddentext transmission rate for $\mathrm{S1}_{\mathrm{corrected}}$ is between 0.045 and 0.2 hiddentext bits per covertext message.

Our construction MESS on the other hand, takes better advantage of D's inherent entropy and does not require expensive error correcting codes. Instead, the stegotext length depends explicitly on the minimum entropy of D: the stegotext is n covertexts long, where $n = \lceil 83/H(D) \rceil$. Therefore for 2^{-80} secrecy/security and at least $1-2^{-80}$ reliability, as long as $H(D) \geq 83$, MESS's stegotext is *only one covertext long*. Moreover, as long as $H(D) \geq 83/4 \approx 21$, MESS's stegotext will be shorter than $\mathrm{S1}_{\mathrm{corrected}}$'s.

6.3 Universal Comparison

The second point of comparison is universality: the schemes' steganographic security when nothing is known about D beyond the assumption that $H(D) \geq 1$. In this case, each scheme must assume the worst possible distribution. For $\mathrm{S1}_{\mathrm{corrected}}$, this means choosing an error correcting code that corrects the worst possible error fraction, $\Delta = 3/8$. For reliability of 2^{-80}, this translates to a stegotext 22 covertexts long and a hiddentext transmission rate of only 0.045 hiddentext bits per covertext for universal $\mathrm{S1}_{\mathrm{corrected}}$. Similarly, for our MESS to be universal with security 2^{-80} and reliability $1 - 2^{-80}$, we must take $n = 83$ and $k = 84$ which translates to a stegotext 83 covertext long and hiddentext transmission rate of only $.012$.

It is readily apparent that $\mathrm{S1}_{\mathrm{corrected}}$ is the superior choice when universality is required (assuming multiple bits are being sent and ignoring the computational cost of error-correction). What we find startling

[6] We reiterate that the definition of stego-system given in [6] and [5] only requires reliability $2/3$, i.e., the probability that each individual hiddentext bit is incorrectly decoded is no more than $1/3$. However, we believe a useful system should have much higher reliability. Therefore, for comparison purposes, we require that both stegosystems be reliable with probability close to 1.

however, is the steep penalty both systems incur to provide universality. If in fact the covertext distribution has high minimum entropy, when $\mathrm{S1}_{\mathrm{corrected}}$ is acting universally, its error correcting code may "over correct" by a factor of 4. In MESS's case, because it is optimized to take advantage of high entropy distributions, the penalty for universality is a much worse factor of 83. Perhaps this "universality penalty" in the complexity-theoretic model of [6] deserves further study.

7 Conclusions

In this work, we have formulated an upper bound on the probability that an adversary will be able to exploit the flaw of Construction 1 in [6] and detect steganographic transmission under this scheme. Despite the presence of this flaw, the bound of Theorem 1 shows that on covertext distributions with sufficiently high minimum entropy, Construction 1 *is* in fact steganographically secret and requires no modification. For lower entropy distributions which cause Construction 1 to become insecure, we have presented an alternate construction MESS which uses repeated sampling to effectively increase the minimum entropy of the covertext distribution. In our scheme the parameter controlling the number of repeated samples drawn requires knowledge of an upper bound on the minimum entropy of the covertext distribution. Comparing it to the revised version of Construction 1 from [5], for covertext distributions with a minimum entropy of at least 21, our scheme MESS takes better advantage of the distribution's inherent entropy. Consequently, for the same error and reliability levels, MESS will have a shorter overall stegotext and a higher hiddentext transmission rate than $\mathrm{S1}_{\mathrm{corrected}}$.

Unfortunately, MESS does not perform as well as the revised Construction 1 when both are operating as universal stegosystems, i.e. with worst case assumptions. In fact, the efficiency of both schemes is greatly reduced when operating universality. Thus, while universality certainly seems, in theory, like a desirable stegosystem property, it is also very apparent that, at least for the stegosystems discussed in this work and likely in general, universality doesn't come cheaply. This is perhaps not so surprising if we consider the physical analog: camouflage. If you want to camouflage an object's presence in an environment, it is much easier to design effective camouflage with *a priori* knowledge of the environment, e.g. urban, desert, jungle, arctic, etc. (and less importantly knowledge of the specific object, e.g. its size, shape, etc.). Therefore, in practice, a scheme like ours which is weakly parameterized by minimum entropy or some other minimal assumption may often be more useful.

References

[1] W. Beyer, editor. *CRC Standard Mathematical Tables and Formulae*. CRC Press, 29 edition, 1991.

[2] C. Cachin. An information-theoretic model for steganography. In *Second Internation Workshop on Information Hiding*, volume 1525 of *Lecture Notes in Computer Science*, pages 306–316, 1998.

[3] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, October 1986.

[4] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[5] N. Hopper, J. Langford, and L. von Ahn. Companion to "provably secure steganography". available from `http://www-2.cs.cmu.edu/~jcl/papers/papers.html`.

[6] N. Hopper, J. Langford, and L. von Ahn. Provably secure steganography. In Moti Yung, editor, *Advances in Cryptology—CRYPTO 2002*, Lecture Notes in Computer Science. Springer-Verlag, 18–22 August 2002. Corrected verstion appears in [7].

[7] N. Hopper, J. Langford, and L. von Ahn. Provably secure steganography. Technical Report CMU-CS-02-149, School of Computer Science, Carnegie Mellon University, 2002.

[8] G. J. Simmons. The prisoners' problem and the subliminal channel. In David Chaum, editor, *Advances in Cryptology: Proceedings of Crypto 83*, pages 51–67. Plenum Press, New York and London, 1984, 22–24 August 1983.