Simple Stateless Steganography

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

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 a flaw in their basic construction. Their proposed fix for this flaw dramatically reduces the efficiency of the construction, because it requires the use of strong error-correcting codes.

Our first contribution is to demonstrate that *the construction that was thought flawed is actually often not*. By carefully analyzing the severity of the flaw in their original construction, we show that it is safe to use under proper conditions—thus eliminating the need for expensive error-correction. Moreover, when such conditions do not hold, we provide an alternative fix for the flaw, which is often more efficient.

In addition, we demonstrate that for memoryless channels, the construction can be used to send multiple bits *statelessly* (maintaining synchronized state between the sender and the recipient, as was proposed for the original construction, is particularly problematic in steganography). We provide tight bounds on the security of such an approach.

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 [7] 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 [7]).

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 recipient. Conversely, the game for the adversary is to distinguish the distribution of transmitted messages from D.

*The work of both authors was partly funded by the National Science Foundation under Grant No. CCR-0311485. The second author's work was also partly funded by a National Physical Science Consortium Fellowship and stipend support from the National Security Agency.

The Flawed Construction. In addition to providing a model, the authors of [7] also present a number of constructions satisfying the definition. The most elementary of them, on which others rely heavily, is called "Construction 1" in [7]. Subsequently, a subtle security flaw was observed. Though the exact effect of the flaw was not analyzed, the flaw was corrected by the authors in [6]. To distinguish between the original and the corrected versions of this construction, we call them $\mathrm{S1}_{\mathrm{original}}$ and $\mathrm{S1}_{\mathrm{corrected}}$, respectively.

The Expensive Fix. $\mathrm{S1}_{\mathrm{original}}$ is an efficient construction: it can transmit one bit of hiddentext for each covertext message, and the decoding and encoding algorithms are very fast (involving just a few applications of a pseudorandom function). Unfortunately, the correction of [6] has a detrimental effect on this efficiency. $\mathrm{S1}_{\mathrm{corrected}}$ requires between 5 and 6 covertext messages (for most distributions D of interest) to transmit one hiddentext bit, and encoding and decoding involves using expensive error-correcting codes.

The reason for such high cost is the high probability of incorrectly decoding an encoded bit. To provide reliability, therefore, $\mathrm{S1}_{\mathrm{corrected}}$ has to first encode the hiddentext in an error-correcting code and then stego-encode the resulting 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 $1/(1 - H_2(1/4)) \approx 5$.

The Stateful Multibit Extension. Both $\mathrm{S1}_{\mathrm{original}}$ and $\mathrm{S1}_{\mathrm{corrected}}$ encode messages one bit at a time. Note that $\mathrm{S1}_{\mathrm{corrected}}$, due to the stretch of the error-correcting codes, must necessarily allow transmissions of hiddentexts longer than one bit.

Encoding of multibit messages is accomplished by having the sender and the recipient maintain a synchronized counter in order to refresh, for each bit, the pseudorandom function key used in the construction. The need for synchrony presents a particular problem in steganography. Unlike in counter-mode symmetric encryption where the counter value can be sent along with the ciphertext in the clear, here this is not possible. Indeed, the counter itself would also have to be steganographically encoded to avoid detection, which brings us back to the original problem of steganographically encoding multibit messages. Thus, strict synchrony between the sender and the recipient is required, and if a single stegotext is dropped, the recipient will fail to decode everything that follows (moreover, standard error-correcting techniques cannot help with this problem).

1.2 Our Contributions

The Fix Is Often Not Needed. Our main result, Theorem 1, demonstrates that the impact of $\mathrm{S1}_{\mathrm{original}}$'s flaw on its security is irrelevant provided D has sufficiently high min-entropy. Specifically, we show that the adversary's advantage in distinguishing transmitted messages from D is at most $2p$ (plus a negligible amount $2\eta_D$), where p is the probability of the most likely element in D. Thus, if D has no elements of high probability (in other words, has high min-entropy), the adversary will be unable to break $\mathrm{S1}_{\mathrm{original}}$. We also show that the bound of $2p$ is tight within a small constant factor.

Taken together these bounds demonstrate that the expensive fix of $\mathrm{S1}_{\mathrm{corrected}}$ is often unnecessary. Thus, our main contribution is to demonstrate that a more efficient construction, once thought flawed, is actually secure under the proper conditions.

[1] The authors of [6] 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.

Cheaper Fix When Needed. Our second contribution is to describe an alternative fix for the flaw when the min-entropy of D is not sufficiently high. We propose a generalization of $\mathrm{S1}_{\mathrm{original}}$ that simply uses D^n (for a small n) instead of D. We call this construction MESS for "Minimum-Entropy-Sensitive Stegosystem" (in particular, for $n = 1$, $\mathrm{S1}_{\mathrm{original}}$ and MESS are the same).

While the technique MESS uses for improving min-entropy is far from novel, proving that the distinguishing advantage of the adversary in this case remains a negligible function of the relevant security parameters is a technical challenge. Because the negligible quantity η_D is distribution-dependent and the distribution changes in MESS as n grows, one cannot simply invoke the result of our main theorem directly.

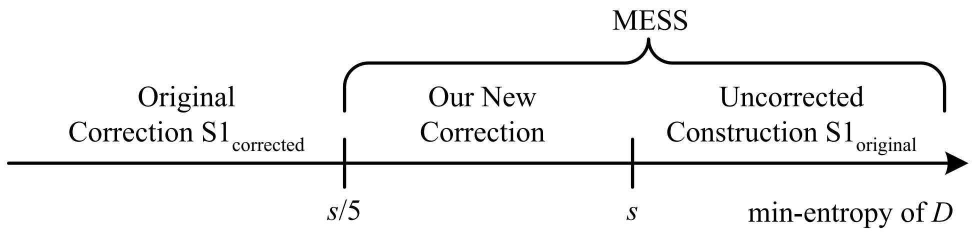

Figure 1: Stegosystems with Highest Data Rate for 2^{-s} Security Level

For comparison purposes we carefully analyze the gains in efficiency that result from using $\mathrm{S1}_{\mathrm{original}}$ or its generalization MESS instead of $\mathrm{S1}_{\mathrm{corrected}}$. In particular, for a security level of 2^{-s}, using MESS results in shorter stegotexts as long as min-entropy of D is at least $(s + 2)/5$. For example, for a common security level of 2^{-80}, MESS has a shorter stegotext whenever min-entropy of D is at least 17. The gains in efficiency of encoding and decoding are even more dramatic. This is because $\mathrm{S1}_{\mathrm{corrected}}$ needs expensive error-correcting codes, while MESS consists simply of repeated sampling from D. Thus, MESS may be beneficial at lower min-entropies as well: even though the data rate will be lower, the computations will be faster.

The Stateless Multibit Extension. Our third significant contribution is to prove that for memoryless covertext distributions D, the scheme $\mathrm{S1}_{\mathrm{original}}$ and its generalization MESS can securely transmit multibit hiddentext messages using bit-by-bit steganographic encoding *without additional state*. In particular, no synchronization between the sender and the recipient is required; therefore, if a portion of the stegotext gets lost in transit, the rest of the message can be correctly recovered.

Specifically, by a non-trivial extension of the techniques used to bound the flaw of $\mathrm{S1}_{\mathrm{original}}$ and to prove the security of MESS, we demonstrate that for a hiddentext message of length l, the distinguishing advantage of the adversary is no more than $6l^2p$ (plus $4l\eta_D$, an amount that remains negligible for reasonable values of l). This bound is also tight within a small constant factor.

Prior to our work, no analysis of such stateless multibit extension was available in the literature. However, Hopper [5] stated that the advantage of the multi-bit construction was loosely quadratic. (Our bound was derived in an attempt to disprove his statement.)

We stress that our result is only for distributions D that are memoryless, i.e., where each covertext message is independent of the history. Proving multibit security of these constructions in the more general case remains an open problem, as far as we know.

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

2.1 Definitions

We reiterate the main definitions and notational conventions from [7] 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 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 when 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 $M(h)$; $SE^M(K, m, h)$ 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

2. 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 $M(h)$; $SD^M(K, c, h)$ returns a hiddentext m.

3. 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 believe it useful for later discussions to elaborate on 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 above definition from [7] considers reliability of $2/3$ sufficient, we will generally consider more reliable stegosystems, since they will likely be of greater practical use.

The Adversary. In this work only passive (i.e., non-malicious) 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]Technically M is made necessary by the fact that we do not assume either complete knowledge of nor 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.

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

3. W can make at most q queries of combined 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 (having messages of fixed length B bits) 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 $\mathbf{InSec}^{\mathrm{SS}}_{\mathcal{S}(\kappa),\mathcal{C}}(t,q,l) \leq \epsilon$.

With respect to the specific constructions discussed herein we need some additional notation which also mirrors that in [7]. 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 $\mathcal{F} : \{0,1\}^\kappa \times \{0,1\}^L \to \{0,1\}$ 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 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 [9].

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 [7] 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 $\mathrm{S1_{original}}$ for brevity, appears in Section 3.2 of [7]. We present it for one-bit messages; as already mentioned, [7] extend it to multibit messages simply by using a synchronized counter. We will address multibit messages in Section 5.

$\mathrm{S1_{original}}$ relies on a *rejection sampler* sub-procedure RS, which appears in Section 2.3 of [7]. This rejection sampler, and consequently $\mathrm{S1_{original}}$, 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, $F : \{0,1\}^B \rightarrow \{0,1\}$. 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 $\mathrm{S1_{original}}$. The specification of RS follows.

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

The stego encoding algorithm SE for $\mathrm{S1_{original}}$ 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 returns the output of RS. The stego decoding or extraction algorithm SD for $\mathrm{S1_{original}}$ takes the key K and a stegotext x and outputs the image of x under F as the hiddentext m.

Procedure $\mathrm{S1_{original}}.SE(K, k, m, h)$:
  $x \leftarrow \mathrm{RS}^{M(h),F(K,\cdot)}(m, k)$
  $h \leftarrow h \| x$
Output: x

Procedure $\mathrm{S1_{original}}.SD(K, x)$:
  $m \leftarrow F(K, x)$
Output: m

From here on we will cease to explicit mention the sampling oracle M and message history h when discussing RS, SE, and SD.

2.3 How and Why $\mathrm{S1_{original}}$ Fails

Corollary 1 in [7] falsely states that $\mathrm{S1_{original}}$ 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

[4]In [7], $k = |K| = \kappa$ is a single security parameter. Here we separate them for precision of discourse.

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 [6] 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, seems 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.

2.4 Revised Construction 1

Hopper, Langford, and von Ahn corrected the flaw of $\mathrm{S1}_{\mathrm{original}}$, described in Section 2.3 of this work, shortly after its publication in [7]. They gave $\mathrm{S1}_{\mathrm{corrected}}$, a revised version that we describe below, in [6], but did so without any analysis of the severity of the flaw. Our main result, presented in Section 3, shows that the flaw can be precisely quantified. Our second result, presented in Section 4, shows that it can, in fact, be made negligible for *any* distribution D.

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 [6] 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.

3 Main Result: 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 $\mathrm{RS}^{D,g}$. We then give a lower bound demonstrating that the upper bound is tight up to a small constant factor.

3.1 Upper Bound

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

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

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 that will yield the proof of Theorem 1.

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. The proof of this Lemma is contained in Appendix A. ☐

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, p. 471]. 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,

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

Finally, combining these two pieces,

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

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 $k \in \mathbb{N}$,*

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

Proof. The proof of this Lemma is contained in Appendix B. □

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. □

3.2 Lower Bound

Theorem 2. *For any p, there exists a probability distribution D with highest-probability element p such that, for any $k > 2$, a bit $b \in \{0,1\}$ and for a randomly chosen predicate $g : D \to \{0,1\}$, the statistical difference between D and $\mathrm{RS}^{D,g}(b,k)$ is at least $p/16$. More precisely,*

$$\sum_{\forall x\in D}\left| \Pr_D[x] - \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,k)\to x]\right| \ge p/8\,.$$

Proof. For lack of space, we only sketch the proof of this theorem. Simply let D consist of $1/(2p)$ elements of probability p each, and $1/(2q)$ elements of probability q, where q is very small. Then one can show that the likelihood that $\mathrm{RS}^{D,g}$ will pick a p-probability element is $p/16$ less than $1/2$. □

4 Generalizing $\mathrm{S1}_{\mathrm{original}}$

We have shown that for D with sufficiently high min-entropy, $\mathrm{S1}_{\mathrm{original}}$ (i.e., Construction 1 of [7]) needs no modification. On the other hand, since p is fixed for any given D, the error of $\mathrm{S1}_{\mathrm{original}}$ is not a negligible function. Thus, when D *lacks* sufficiently high min-entropy, $\mathrm{S1}_{\mathrm{original}}$ in its current form is *insecure*. This brings us to our second 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 $\mathrm{S1}_{\text{original}}$ 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 the well known technique of 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 n 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 value of n depends on $H(D)$ and is fixed for a given D. Our proposed stegosystem MESS is the same as $\mathrm{S1}_{\text{original}}$ except for a few minor syntactic changes necessary to accommodate its use of RS-HE instead of RS.

Thus, MESS has three security parameters: $\kappa = |K|$, k and n, which are, respectively, the length of the pseudorandom predicate key, the number of attempts made by RS-HE, and the number of draws from D that are concatenated and given to the pseudorandom predicate. Let MESS(κ, k, n) denote our new system instantiated with these parameters. For a formal description of MESS see Appendix C.

4.2 Proof of Correctness

The proof of $\mathrm{S1}_{\text{original}}$ given in [7] only attempted to show security with respect to adversaries making a single 1-bit query. In this section, we will initially do the same, because multi-bit security follows from 1-bit security by use of a synchronized counter as in [7]. Later we will show that for the special case of memoryless channels, our techniques can be adapted to prove stateless multibit security.

The proof that MESS is 1-bit steganographically secure follows (although not immediately) 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.

Theorem 3. *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 for any $0 < \delta < 1/2$,*

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

Proof. As already stated, the hard part is to bound $\eta(D^{(n)}, k)$. We actually bound a closely related value $\zeta(D^{(n)}, k)$. This relies on two lemmas: 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.

The detailed proof of this theorem (including the lemmas) is contained in Appendix D. □

4.3 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 4. *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 $0 < \delta < \frac{1}{2}$,*

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

Proof. The proof of this Theorem is contained in Appendix E. □

4.4 MESS Parameter Choices and Efficiency

Given covertext distribution D with min-entropy $H(D)$, for MESS to operate with 2^{-s} security and a corresponding reliability of at least $1-2^{-s}$ (for $s \geq 13$), it suffices to take $n = \lceil (s+2)/H(D) \rceil$, $k = s+6$, and κ such that for the chosen PRF family $\mathcal{F}$, $\mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k) \leq 2^{-s-3}$ (the derivation of these parameter values can be found in Appendix F). The stegotext is just n covertexts long.

In Appendix G, we show that to achieve reasonable reliability, $\mathrm{S1}_{\mathrm{corrected}}$ needs to send more than 5 covertexts for each hiddentext bit (more for distributions with really low min-entropy). Thus, if $H(D) \geq (s+2)/5$, MESS sends fewer covertexts than $\mathrm{S1}_{\mathrm{corrected}}$, and if $H(D) \geq (s+2)$, MESS sends only a single covertext, effectively reducing to $\mathrm{S1}_{\mathrm{original}}$. Moreover, MESS requires no computationally expensive error-correction.

5 Stateless Multibit Extension of MESS

Having addressed the security flaw of $\mathrm{S1}_{\mathrm{original}}$ for 1-bit hiddentexts by demonstrating the security of the more general construction MESS in the 1-bit case, we now consider secure transmission of multibit hiddentext messages. As previously mentioned, a secure stateful multibit version of MESS can be obtained, as was done in [7]. Namely, the sender and recipient maintain a synchronized counter c and do straightforward bit-by-bit stego-encoding with MESS by providing c as an additional input to the PRF. The counter essentially serves to refresh the pseudorandom function key, thereby making each successive hiddentext bit as secure as the first. However, as we will show next, if the covertext message distribution D is memoryless, we can achieve secure stateless multibit steganographic encodings by directly doing bit-by-bit stego-encoding using MESS, thus eliminating the need for a synchronized counter.

Theorem 5. *Let D be a memoryless covertext message distribution, and let p be the probability of the most likely element of D ($p = 2^{-H(D)}$). Then for a total of $l \geq 1$ hiddentext bits transmitted (chosen by the adaptive warden)*

$$\mathbf{InSec}^{\mathrm{SS}}_{MESS(\kappa,k,n),D}(t,l,l) \leq 6l^2p^n + l\left(\left(\frac{1}{2}+\delta\right)^k + \mathrm{e}^{-\lfloor \frac{1}{p^n} \rfloor 2\delta^2}\right) + \mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(t+O(lk), lk)\,.$$

Proof. The proof makes use of two key lemmas for memoryless distributions D. The first, Lemma 7, shows that the advantage of any adversary *adaptively* asking for the stego-encoding of a total of l bits of hiddentext can be bounded by the advantage of a *non-adaptive* adversary that asks $2l$-bit hiddentext queries of the form 1^l0^l. The second, Lemma 8, shows that a string $x = x_1x_2 \ldots x_{2l}$ that contains no repeated elements is no less likely to occur as a stego encoding of 1^l0^l than as a random draw from D^{2l} (provided sampler $\mathrm{RS}^{D,g}$ is allowed to make as many draws as needed, i.e. $k = \infty$). The rest follows from (1) using Lemma 8 to argue that the statistical difference is no more than three times the probability of the existence of a collision among $2l$ elements chosen from D; (2) bounding the probability of such a collision and (3) dealing with finite k.

A more detailed proof of this Theorem (including the lemmas) is contained in Appendix H. □

This bound is also nearly optimal: we show that the adversary who asks $l/2$ 1-queries followed by $l/2$ 0-queries can distinguish with probability roughly $l^2p^n/4$.

Theorem 6. *For any p, there exists a probability distribution D with highest-probability element p such that, for a randomly chosen predicate $g : D \to \{0,1\}$, the statistical difference between D^l and* $\mathrm{RS}^{D,g}(1^{l/2}0^{l/2}, \infty)$

is greater than one half the probability of obtaining a collision among l draws from D.

$$\sum_{\forall x \in D^l} \left| \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(m,k) \to x] - \Pr_{D^l}[x] \right| \geq \frac{pl^2}{4} - \left(\frac{pl^2}{4}\right)^2 .$$

Proof. The proof of this Theorem is contained in Appendix I. It is obtained by comparing the probability of a collision between an answer to a 1-query and an answer to a 0-query, which is 0 for $\mathrm{RS}^{D,g}$ and non-zero for D^l. □

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. Private Communication.

[6] 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`.

[7] 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 [8].

[8] 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.

[9] 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.

A Proof of Lemma 1

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. $\square$

Remark 1. It can be seen from (1) 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 [6]. Indeed, the proposed fix in [6] 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.

B Proof of Lemma 3

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

$$\begin{aligned}
&\sum_{\forall x\in D}\left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,\infty) \to x] - \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(b,k) \to x]\right| && (2)\\
&= \sum_{\forall x\in D}\frac{\Pr[x]}{2^{|D|}}\left|1 + \sum_{S\subseteq D:x\in S}\frac{1}{\alpha_S} - \sum_{S\subseteq D:x\in S}\frac{\beta_S^k - 1}{\beta_S - 1} - \sum_{S\subseteq D:x\notin S}\beta_S^{k-1}\right| && (3)\\
&= \sum_{\forall x\in D}\frac{\Pr[x]}{2^{|D|}}\left|1 + \sum_{S\subseteq D:x\in S}\frac{1}{\alpha_S} - \sum_{S\subseteq D:x\in S}\frac{1-\beta_S^k}{\alpha_S} - \sum_{S\subseteq D:x\in S}\alpha_S^{k-1}\right| && (4)\\
&= \sum_{\forall x\in D}\frac{\Pr[x]}{2^{|D|}}\left|\sum_{S\subsetneq D:x\in S}\frac{\beta_S^k - \alpha_S^k}{\alpha_S}\right| && (5)\\
&= 2\sum_{x\in D:|\cdot|\geq 0}\frac{\Pr[x]}{2^{|D|}}\sum_{S\subsetneq D:x\in S}\frac{\beta_S^k - \alpha_S^k}{\alpha_S} && (6)\\
&\leq \frac{1}{2^{|D|-1}}\sum_{\forall x\in D}\Pr[x]\sum_{S\subsetneq D:x\in S}\frac{\beta_S^k}{\alpha_S} && (7)\\
&= \frac{1}{2^{|D|-1}}\sum_{S\subsetneq D:S\neq\emptyset}\frac{\beta_S^k}{\alpha_S}\sum_{\forall x\in S}\Pr[x] && (8)
\end{aligned}$$

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

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

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. □

C Formal Description of MESS

C.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$^{D,F}(y, \mathit{count}, n)$:
  $i = 0$
  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 $\mathit{count} = i$
Output: x

The only differences between the stego-encoding algorithms for 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.

C.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 C.1 with the stipulation that now $F_K : D^{(n)} \to \{0, 1\}$.

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 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 $D^{(n)}$.

D Proof of Theorem 3

Before proving Theorem 3 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 Lemma 6 (in Appendix F), ζ is exactly the failure probability of the rejection sampler RS-HE^{D,g}.

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 $k \in \mathbb{N}$, $\zeta(D, k) \leq \zeta(U_D, k)$*

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 3. 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 C.2, such an $M^{(n)}$ can be easily constructed from M and, in fact, RS-HE$^{M(\cdot),F}(b,k)$ is equivalent to RS$^{M(\cdot),F}(b,k)$ for the same predicate F. Thus applying Theorem 1 gives,

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

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

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

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

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

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

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

□

E Proof of Theorem 4

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 4. The reliability of MESS(κ, k, n) is simply one minus the encoding error introduced by 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 3 it was argued that RS-HE$^{D,F_K}(\cdot, k, n)$ and RS$^{D^{(n)},F_K}(\cdot, k)$ are equivalent (see also Remark 2 of Appendix C.2). So, by Lemma 6 and the definition of pseudorandom function insecurity, the encoding error introduced by 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 3 and subtracting from one gives the indicated lower bound for the reliability. □

F Parameter Derivation and Running Time for MESS

Given covertext distribution D with min-entropy $H(D) > 1$, for MESS to operate with 2^{-s} security and a corresponding reliability of at least $1 - 2^{-s}$, what values of the parameters κ, k, and n suffice? First, we take $n \geq (s+2)/H(D)$, so that $2p^n < 2^{-s-1}$. Then we take $k = s + 6$. If we set $\delta = 1/(4(s+4))$, then the term $2(1/2 + \delta)^k = 2(1/2)^k(1 + 2\delta)^k \leq 2^{-k+1}(1 + 1/k)^k < 2^{-k+3} = 2^{-s-3}$. In order for the third term to be at most 2^{-s-3}, we need $\lfloor 1/p^n \rfloor 2\delta^2 \log_2 e \geq s + 4$. Substituting 2^{s+2} for $1/p^n$ and $1/(4(s+4))$ for δ, we get that we need $log_2 e 2^{s+2} \geq 8(s+4)^3$, which holds as long as $s \geq 13$ (insecurity greater than 2^{-13} is not acceptable in most applications, anyway, so this is not really a restriction).

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^{-s-3}. These same parameter choices will also provide the desired reliability level.

Note that the value of k specified here is the *maximum* number of attempts RS-HE makes, but the *expected* number of attempts is just 2.

For each hiddentext bit, the stego-encoder for MESS essentially just draws, on average, $2n$ samples from the covertext distribution D and thus evaluates, on average, 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. Thus, the running time of our decoder is essentially one PRF evaluation, and the average running time of our encoder is about twice that. The stegotext length is clearly just n covertexts long.

Final Values: To obtain 2^{-s} security and a corresponding reliability of at least $1 - 2^{-s}$ for MESS, as long as $s \geq 13$, it suffices to take $n \geq \lceil (s+2)/H(D) \rceil$, $k = s + 6$, and κ such that for the chosen PRF family $\mathcal{F}$, $\mathbf{InSec}^{\mathrm{PRF}}_{\mathcal{F}(\kappa)}(O(nk), k) \leq 2^{-s-3}$. Thus for example, for 2^{-80} security and $1 - 2^{-80}$ reliability, if $H(D) \geq 82$ then *MESS's stegotext is only one covertext long* (that is, MESS simplifies to $\mathrm{S1}_{\mathrm{original}}$).

G Parameter Derivation for $\mathrm{S1}_{\mathrm{corrected}}$.

Here we demonstrate that for secure and reliable transmission, $\mathrm{S1}_{\mathrm{corrected}}$ needs to send $1 - H_2(1/2 - 1/4(1-p))$ covertexts per hiddentext bit, where H_2 is the binary entropy. This value is between 5 and 6 for reasonable p ($p < .05$).

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

[6]We reiterate that the definition of stego-system given in [7] and [6] 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.

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

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

H Proof of Theorem 5

As we did in the proof of Theorem 3 we will develop intermediate results for the simpler case of $\mathrm{S1}_{\mathrm{original}}$ and then generalize the results for MESS. In the following let $\mathrm{RS}^{D,g}(1^l0^l,\infty) \to x$ denote the event that the rejection sampler on bit-by-bit input 1^l0^l outputs $x = x_1x_2\ldots x_{2l}$ where each $x_i \in D$ and $g : D \to \{0,1\}$ is a randomly chosen predicate.

Lemma 7. *The advantage of any adversary W' that* adaptively *asks an oracle for MESS(κ, k, n) for the bit-by-bit stego-encoding of a total of l bits of hiddentext is bound by the advantage of a* non-adaptive *adversary that asks the same oracle for the bit-by-bit stego-encoding of the $2l$-bit hiddentexts 1^l0^l.*

Proof. Suppose W' adaptively asks for the bit-by-bit encoding of $m = m_1m_2\ldots m_l$, $m_i \in \{0,1\}$. W simply first asks its corresponding stego-encoding oracle for the bit by bit encoding of 1^l0^l. Then W just uses the encoding of the ith zero or the ith one that it received to answer W''s ith adaptively chosen query. Since the draws from D are independent, the distribution of stego-encodings that W' receives from W is identical to that it would have received directly. □

Lemma 8. *For all $x = x_1x_2\cdots x_{2l} \in D^{2l}$ such that $\forall i,j\ x_i \neq x_j$, that is for any strings of $2l$ elements from D which does not contain a repeated element,*

$$\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty) \to x] \geq \Pr_{D^{2l}}[x]\,.$$

The proof of Lemma 8 makes use of the following fact.

Proposition 2. *For any set of n non-negative real numbers $a_1, a_2, \ldots, a_n$ and $l > 1$,*

$$\frac{1}{n}\sum_{i=1}^{n} a_i^l \geq \left(\frac{\sum_{i=1}^{n} a_i}{n}\right)^l.$$

Proof of Lemma 8. Combining the fact that successive draws from D are independent, i.e. $\Pr_{D^{2l}}[x] = \Pr_D[x_1]\Pr_D[x_2]\cdots\Pr_D[x_{2l}]$, with line (1) from the proof of Lemma 1 gives,

$$\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty) \to x_1x_2\cdots x_{2l} \text{ s.t. } \forall i,j\ x_i \neq x_j] \tag{13}$$

$$= \frac{\Pr_{D^{2l}}[x]}{2^{|D|}} \sum_{S\subset D: x_1,x_2,\cdots,x_l\in S \wedge x_{l+1},\cdots,x_{2l}\notin S} \frac{1}{\alpha_S^l\beta_S^l} \tag{14}$$

$$= \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \frac{1}{2^{|D|-2l}} \sum_S \frac{1}{\alpha_S^l(1-\alpha_S)^l} \tag{15}$$

$$\geq \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \left(\frac{\sum_S \frac{1}{\alpha_S(1-\alpha_S)}}{2^{|D|-2l}}\right)^l \tag{16}$$

$$\geq \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \left(\frac{2^{|D|-2l}}{\sum_S \alpha_S(1-\alpha_S)}\right)^l \tag{17}$$

$$\geq \quad \frac{\Pr_{D^{2l}}[x]}{2^{2l}}\left(\frac{2^{|D|-2l}}{\sum_{i=1}^{2^{|D|-2l}} 1/4}\right)^{l} \tag{18}$$

$$= \quad \frac{\Pr_{D^{2l}}[x]4^{l}}{2^{2l}} \tag{19}$$

$$= \quad \Pr_{D^{2l}}[x]\,. \tag{20}$$

Line (16) follows from Proposition 2, line (17) from Proposition 1, and line (18) from the fact that $\max_{0<\alpha<1}\alpha(1-\alpha)=1/4$, and the remaining lines follow from algebra. □

Corollary 1 (To Lemma 8: Non-collision Statistical Difference). *The statistical difference between $\mathrm{RS}^{D,g}(1^l 0^l,\infty)$ and D^{2l} for elements $x=x_1x_2\cdots x_{2l}\in D^{2l}$ such that no value x_i is repeated, i.e. $\forall i,j$ such that $1\leq i\neq j\leq l$ $x_i\neq x_j$, is less than probability of drawing an element x from D^{2l} containing at least one repeated element. Namely,*

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\neq x_j}\left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]-\Pr_{D^{2l}}[x]\right|\leq\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{D^{2l}}[x]\,.$$

Proof. The proof follows directly from Lemma 8 by opening the absolute value signs on the statistical difference and replacing the probability of no collisions by the probability of one minus the probability of a collision in each of the distributions. That is,

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\neq x_j}\left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]-\Pr_{D^{2l}}[x]\right|=$$

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\neq x_j}\left(\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]-\Pr_{D^{2l}}[x]\right)=$$

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\neq x_j}\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]-$$

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\neq x_j}\Pr_{D^{2l}}[x]$$

$$=\left(1-\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]\right)-$$

$$\left(1-\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{D^{2l}}[x]\right)$$

$$=\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{D^{2l}}[x]-$$

$$\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\rightarrow x]$$

$$\leq\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j}\Pr_{D^{2l}}[x]\,.$$

□

Corollary 2 (To Lemma 8). *The probability that* $\mathrm{RS}^{D,g}(1^l0^l,\infty)$ *outputs an element* $x = x_1x_2\cdots x_{2l} \in D^{2l}$ *such that at least one value* x_i *is repeated, i.e.* $\exists i,j$ *such that* $1 \le i \ne j \le l$ *and* $x_i = x_j$*, is less than or equal to the probability of drawing such an* x *from* D^{2l} *directly.*

Proof. Since by Lemma 8, for every string x of $2l$ unique elements from D, $\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty) \to x] \ge \Pr_{D^{2l}}[x]$,

$$\begin{aligned}
\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i=x_j} \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x] = \\
&1-\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\ne x_j} \Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x]\\
\le\ &1-\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\ne x_j} \Pr_{D^{2l}}[x]\\
=\ &\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \Pr_{D^{2l}}[x]\,.
\end{aligned}$$

□

Lemma 9. *Let* D *be any memoryless discrete probability distribution and* p *be the probability of the most likely event in* D*. Then for the hiddentext bit string* 1^l0^l *for any* $1 \le l$ *and a randomly chosen predicate* $g : D \to \{0,1\}$*, the statistical difference between* D^{2l} *and* $\mathrm{RS}^{D,g}(1^l0^l,\infty)$ *is at most* $6l^2p$*. More precisely,*

$$\sum_{\forall x\in D^{2l}} \left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x] - \Pr_{D^{2l}}[x]\right| \le 6l^2p\,.$$

Proof of Lemma 9. Splitting the statistical difference into the collision and non-collision components, then applying Corollary 1 and the triangle inequality, next applying Corollary 2, and finally upper bounding the probability of collisions on l draws from D by $2l^2p$ (derived using counting and the union bound) gives the stated results. More precisely,

$$\begin{aligned}
\sum_{\forall x\in D^{2l}} \left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x] - \Pr_{D^{2l}}[x]\right| = \\
&\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\forall i,j x_i\ne x_j} \left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x] - \Pr_{D^{2l}}[x]\right| + \\
&\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x] - \Pr_{D^{2l}}[x]\right|\\
\le\ &\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \Pr_{D^{2l}}[x] + \\
&\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \left|\Pr_{g\in U(B,1),D}[\mathrm{RS}^{D,g}(1^l0^l,\infty)\to x]\right| + \\
&\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \left|\Pr_{D^{2l}}[x]\right|\\
\le\ &3\sum_{x=x_1x_2\cdots x_{2l}\in D^{2l}|\exists i,j x_i= x_j} \Pr_{D^{2l}}[x]\\
\le\ &6k^2p\,.
\end{aligned}$$

□

Lemma 10. *For a fixed $k, l \in \mathbb{N}$,*

$$\sum_{\forall x \in D^{2l}} \left| \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(1^l 0^l, \infty) \to x] - \Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(1^l 0^l, k) \to x] \right| \leq 4l\eta(D,k)$$

Proof. This sum captures the difference in probabilities between the rejection sampler in the infinite and finite cases. The element $x = x_1 x_2 \ldots x_{2l}$ will be output in the infinite case, but not in the finite case, whenever at least one x_i is output by RS after more than k attempts. Thus, because D is memoryless, taking the union over the $2l$ components with the probability that each element needed more than k draws from Lemma 3 for the 1-bit case, the stated bound follows directly. □

Proof (sketch) of Theorem 5. The structure of the proof is similar to that of Theorem 3. The proof follows by first inserting positive and negative $\Pr_{g \in U(B,1),D}[\mathrm{RS}^{D,g}(1^l 0^l, \infty) \to x]$ inside the absolute value signs, applying the triangle inequality, and then using Lemmas 9 and 10 with D^n in place of D to account for the repeated sampling by MESS. Then $\eta(D^n, k)$ is bound using Lemma 5 as in the proof of Theorem 3. Finally, adjusting for the advantage due to a pseudorandom F gives the desired result. □

I Proof of Theorem 6

Proof. Assume for simplicity that l is even and let D be the uniform distribution: D has $1/p$ elements of probability p each. Let $x_1 \ldots x_l$ be the elements drawn. Simply consider the probability that there exists a collision between x_i and x_j, $1 \leq i \leq l/2 < j \leq l$. It is 0 in the case of $\mathrm{RS}^{D,g}(1^{l/2} 0^{l/2}, \infty)$.

Now in the case of D^l, first think of choosing all of the elements first and then randomly assigning them to either half. If there is a collision among the l elements drawn, then the probability that colliding elements end up in different halves at least $\frac{l}{2(l-1)}$. Next, we lower bound the probability of collisions among and l element draw from D in general by upper bounding the probability of non-collisions as follows,

$$\begin{aligned}
\sum_{x = x_1 x_2 \cdots x_l \in D^l | \forall i \neq j \ x_i \neq x_j} \Pr_{D^l}[x] &= (1-p)(1-2p)\cdots(1-(l-1)p) && (21) \\
&\leq \mathrm{e}^{-p-2p-\cdots-(l-1)p} && (22) \\
&= \mathrm{e}^{-pl(l-1)/2} && (23) \\
&\leq 1 - pl(l-1)/2 + (pl(l-1)/2)^2/2\,. && (24)
\end{aligned}$$

Line (22) and Line (24) follow from the Taylor series expansion of e^{-x} which gives $(1-x) \leq \mathrm{e}^{-x} \leq 1 - x + x^2/2$. Thus the probability of collisions among the l elements drawn from D is,

$$\begin{aligned}
\sum_{x = x_1 x_2 \cdots x_l \in D^l | \exists 1 \leq i \leq l/2 < j \leq l i : x_i = x_j} \Pr_{D^{2l}}[x] &= 1 - \sum_{x = x_1 x_2 \cdots x_l \in D^l | \forall i \neq j \ x_i \neq x_j} \Pr_{D^2}[x] \\
&\geq 1 - \left(1 - \frac{pl(l-1)}{2} + \frac{(pl(l-1)/2)^2}{2}\right) \\
&= \frac{pl(l-1)}{2} - \frac{(pl(l-1)/2)^2}{2}\,.
\end{aligned}$$

Multiplying this by $\frac{l}{2(l-1)}$ from above gives the lower bound of $pl^2/4 - (pl^2/4)^2$. □