

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Thesis

**SIMPLE, STATELESS STEGANOGRAPHY
FOR MEMORYLESS CHANNELS**

by

SCOTT W. RUSSELL

B.A., Hamilton College, 1992
M.S., Northwestern University, 1993

Submitted in partial fulfillment of the
requirements for the degree of
Master of Arts

2004

Approved by

First Reader

Leonid Reyzin, Ph.D.
Assistant Professor of Computer Science

Second Reader

Gene Itkis, Ph.D.
Assistant Professor of Computer Science

Acknowledgments

Many thanks to Professor Leonid Reyzin for his time, energy, inspiration, and invaluable assistance with this work. The author would also like to thank Professor Gene Itkis for his help and insights on this and other unrelated work. Additional thanks to the members of the Applied Cryptography and e-Security (ACeS) group for their comments and questions regarding this work. Finally, thanks to the anonymous reviewers of various conference versions of this work.

**SIMPLE, STATELESS STEGANOGRAPHY
FOR MEMORYLESS CHANNELS**

SCOTT W. RUSSELL

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 such stegosystems. Subsequently, a flaw was discovered in the security analysis of 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 to ensure high reliability.

The contributions of this work are three-fold. First, it demonstrates that *the construction that was thought flawed is actually often not*. By carefully analyzing the severity of the flaw in the original construction, this work shows that it is safe to use under the proper conditions—thus eliminating the need for expensive error-correction. Second, when such conditions do not hold, *an alternative and often more efficient generalized construction addressing the flaw is described*. Lastly, for memoryless channels, *the newly proposed construction can be used to send multiple secret bits statelessly*. Stateless constructions are particularly important in steganography where communication attempting to resynchronize the shared state is likely to alert the adversary.

Contents

0	List of Abbreviations	vii
1	Introduction	1
1.1	Background	1
1.2	Contributions of This Work	3
1.3	Subsequent Work	5
2	Background: Work of Hopper, Langford, and von Ahn	6
2.1	Definitions	6
2.2	Problematic Construction	10
2.3	Problem With the Construction	12
2.4	Revised Construction	13
3	Bounding the Magnitude of the Problem	14
3.1	Insecurity Upper Bound	14
3.2	Supporting Results	15
4	MESS: A Secure Generalization of the Problematic Construction	18
4.1	Construction Description	18
4.2	MESS Security	19
4.3	MESS Reliability	21
4.4	MESS Parameter Choices and Efficiency	23
5	Stateless, Multibit Extension of MESS for Memoryless Distributions	23
5.1	Supporting Results for Multibit Insecurity Upper Bound	25
5.2	Optimality of Multibit Bound	27

A Proof of Lemma 1	29
B Proof of Lemma 3	30
C Formal Description of MESS	31
C.1 For Memoryless Distributions	31
C.2 For General Distributions	32
D Proofs of Lemma 4 and 5, and Theorem 2	33
E MESS Parameter Derivation and Running Time	37
F Revised Construction 1 Parameter Derivation	38
G Proofs of Lemmas 7 and 8, Corollaries 1 and 2, and Lemmas 9 and 10	38

0 List of Abbreviations

(\cdot) denotes a function of a single input

$\|$ denotes concatenation

$|$ also used to denote concatenation

$\lceil \cdot \rceil$ denotes the ceiling function

$\lfloor \cdot \rfloor$ denotes the floor function

\leftarrow gets the value

\rightarrow outputs the value

$||$ denotes either absolute value or set cardinality

\approx approximately equal to, asymptotically close to

$\stackrel{def}{=}$ defined to be equal to

A pseudorandom function adversary

\mathcal{A} class of pseudorandom function adversaries

$\mathbf{Adv}_{\mathcal{F}(\kappa, n)}^{\text{PRF}}(A)$ advantage of pseudorandom function adversary A against pseudorandom function family \mathcal{F} instantiated with parameters κ and n

$\mathbf{Adv}_{S(\kappa), \mathcal{C}}^{\text{SS}}(W)$ advantage of stegosystem adversary W against stegosystem S instantiated with parameters κ and operating over channel \mathcal{C}

α_S weight of elements in subset S

α_G weight of elements mapping to 1 under predicate G

b a single bit

B fixed length in bits of elements in support of distribution D

β_S weight of elements in complement of subset S ($\beta_S = 1 - \alpha_S$)

β_G weight of elements mapping to 0 under predicate G ($\beta_G = 1 - \alpha_G$)

\mathcal{C} unidirectional communication channel, modeled as a distribution of bit sequences

\mathcal{C}_h^B distribution of sequences of B -bit blocks conditioned on history h

C channel capacity

D distribution of bit sequences, or alternatively distribution of covert messages

D^n distribution of sequences of n consecutive elements drawn from D

Δ stego-encoding error rate

e base of the natural logarithm

ε a small positive quantity

$\eta(D, k)$ average of k th power subset weights α_S^k for distribution D (excluding full set)

η_D abbreviation for $\eta(D, 1)$

\mathcal{F} pseudorandom function family

F_K pseudorandom function with key K

$F_{K,n}$ pseudorandom function with key K , and input domain parameterized by n

G random function over fixed length bit strings

h history of previously drawn elements from support of distribution

H_∞ minimum entropy

H_2 binary entropy

$\text{InSec}_{\mathcal{F}(\kappa,n)}^{\text{PRF}}$ insecurity of pseudorandom function family \mathcal{F} instantiated with parameters κ and n

$\text{InSec}_{S(\kappa),\mathcal{C}}^{\text{SS}}$ insecurity of stegosystem S instantiated with parameter κ and operating over channel \mathcal{C}

κ security parameter and length of key K ($\kappa = |K|$)

k maximum number of sampling rounds attempted by rejection sampler

K shared, secret key having length $|K| = \kappa$

ℓ error-correcting code stretch factor

l length in bits

m hiddentext message to be stego-encoded

M sampling oracle for distribution D

M^n sampling oracle for distribution D^n

$\max_{x \in D} \{ \}$ maximum of quantity in braces taken over all elements x in the support of distribution D

MESS acronym for “Minimum-Entropy-Sensitive Stegosystem”

$\min_{x \in D} \{ \}$ minimum of quantity in braces taken over all elements x in the support of distribution D

min-entropy abbreviation for minimum entropy

n parameter for MESS, stegotext block length

\mathbb{N} the set of natural numbers

$O(\cdot)$ asymptotic upper bound

$\mathcal{O}_0(m)$ oracle which on input message m , outputs (an appropriate number of) elements sampled directly from the distribution D

$\mathcal{O}_1(m)$ oracle which on input message m , outputs the stego-encoding of m using the specified stegosystem

\mathcal{O}_b oracle corresponding to bit $b \in \{0, 1\}$

p maximal probability in distribution D , or an upper bound for it ($p = 2^{-H_\infty(D)}$)

PRF pseudorandom function

q number of adversary queries

r string of random bits

$\mathbf{Rel}_{S(\kappa), \mathcal{C}}(l)$ reliability of stegosystem S instantiated with parameter κ over channel \mathcal{C} when sending l -bit hiddentext messages

RS rejection sampler subroutine

$\mathbf{RS}^{D,G}$ denotes oracle access by the rejection sampler subroutine to the distribution D and function G (similar superscript interpretation for other algorithms and functionalities)

RS-HE high entropy version of rejection sampler subroutine

S subset of elements from support of a distribution, or alternatively a stegosystem

$S1_{\text{orig}}$ abbreviation for “Construction 1” in [7]

$S1_{\text{corr}}$ abbreviation for revised version of “Construction 1” in [6]

SD stegosystem decoding algorithm

SE stegosystem encoding algorithm

stegosystem abbreviation for steganographic system

stego-decoder abbreviation for steganographic decoding algorithm

stego-encoder abbreviation for steganographic encoding algorithm

t adversary running time

$U(B)$ uniform distribution on strings of B -bits

$U(B, w)$ uniform distribution on function from B -bit strings to w -bit strings

w parameter for MESS, rate specifying number of hiddentext bits encoded per stegotext block

W stegosystem adversary or warden

\mathcal{W} class of stegosystem adversaries

x, y generic elements from support of distribution

X^Y denotes oracle access by algorithm X to functionality Y

$\zeta(D, k)$ average of k th power subset weights α_S^k for distribution D (including full set)

ζ_D abbreviation for $\zeta(D, 1)$

1 Introduction

1.1 Background

The goal of Steganography is for one party, the sender, to communicate a secret message to a second party, the recipient, using observed, innocuous-looking messages. In other words, steganography involves producing *stegotext* messages by subtle alteration to or selection of public *coverttext* messages in order to convey a secret *hiddentext* message. This must be done in such a way that no observer other than the intended recipient is able to distinguish stegotexts from coverttexts. In CRYPTO 2002, Hopper, Langford and von Ahn [7] offered the first rigorous complexity-theoretic formulation of steganography. They formally define *steganographic secrecy* (i.e. security) of a stegosystem as the inability of a polynomial-time adversary to distinguish between the distribution of stegotext output by the stegosystem and the distribution of coverttext. 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 coverttexts that the adversary expects to see. All parties are allowed to draw from D ; the game for the sender is to alter D imperceptibly from the perspective of 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 . As Hopper et al. [6] point out, the underlying channel distribution may well be very complex and not easily described (e.g., human email traffic or images of various scenes). Thus, to obtain a stegosystem of widest applicability, it is desirable to assume as little knowledge about the channel as possible.

This work considers unidirectional channels from the sender to the recipient. It is not assumed that the channel is represented by a polynomial-time sampler that starts

with uniformly distributed random coins (as do [10] and one of the constructions in [13]). Furthermore, it is not even assumed that there is an unbiased function on the channel known to the sender and the recipient (see some constructions in [2, 6, 13]). In many setting, both of these assumptions are unrealistic, and thus limit the applicability of constructions that rely on them.

Basic (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 [9]. Though the exact effect of the flaw was not analyzed, the flaw was corrected by Hopper, et al. in [6]. To distinguish between the original and the corrected versions of this construction, let them be denoted $S1_{\text{orig}}$ and $S1_{\text{corr}}$, respectively.

Expensive Fix. $S1_{\text{orig}}$ is an efficient construction: it can transmit one bit of hiddentext for each covert 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. $S1_{\text{corr}}$ requires between 5 and 6 covert messages (for most distributions D of interest) to transmit one hiddentext bit, and encoding and decoding involves the use of expensive error-correcting codes.

The error-correcting codes are used to ensure high reliability, that is a high probability of correctly stego-decoding each stego- encoded bit. The error correcting code is applied to the hiddentext, and then the resulting codewords are stego-encoded¹. The high rate of error in stego-encoding (between 1/4 and 3/8, depending on D) provides an easy upper

¹The 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.

bound on the rate of the error-correcting code used, and thus a lower bound on the stretch factor, which must be $1/(1 - H_2(1/4)) \approx 5$.

Stateful Multibit Extension. Both $S1_{\text{orig}}$ and $S1_{\text{corr}}$ encode messages one bit at a time. Note that $S1_{\text{corr}}$, 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 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 stego-encoded to avoid detection. But, this reduces the problem of sending the counter to the problem the counter is supposed to help solve, namely steganographically encoding multibit messages. Thus, strict synchrony between the sender and the recipient is required. Otherwise, if a single stegotext is dropped, the recipient may fail to decode everything that follows. Achieving such synchrony when it is not inherently present may be messy and difficult and falls outside the scope of this work. The construction presented herein will not suffer from this problem.

1.2 Contributions of This Work

Fix Often Not Needed. The main result of this work, Theorem 1, demonstrates that the impact of the flaw in the security analysis of $S1_{\text{orig}}$ is irrelevant provided D has sufficiently high min-entropy. Specifically, for a coverttext distribution D with largest probability p , this work shows that for the distribution of messages output by $S1_{\text{orig}}$ and those taken directly from D , the distinguishing advantage of the adversary is at most $2p$ (plus a negligible amount). Thus, if D has no elements of high probability (in other words, has high min-

entropy), the adversary will be unable to break $S1_{\text{orig}}$. Consequently, the expensive fix of $S1_{\text{corr}}$ is often unnecessary. Thus, the main contribution of this work is to demonstrate that a more efficient construction, once thought flawed, is actually secure under the proper conditions.

Cheaper Fix When Needed. The second contribution is a construction called MESS that provides an alternative, often more efficient fix for the flaw when the min-entropy of D is not sufficiently high. MESS which stand for “Minimum-Entropy-Sensitive Stegosystem” is a generalization of $S1_{\text{orig}}$ that simply uses n successive draws rather than a single draw from D to encode each bit of hiddentext. MESS, like $S1_{\text{orig}}$, requires the sender and recipient to have knowledge of a lower bound on the channel’s min-entropy, and the ability to sample from the channel. The recipient needs only to be able to separate one message from the next, and agree with the sender on the min-entropy bound; it needs no other access to the channel. For a parameter setting of $n = 1$, $S1_{\text{orig}}$ and MESS are the same.

While the repeated sampling technique MESS uses to improve 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 presents a technical challenge. Because the negligible quantity is now distribution-dependent and the distribution changes in MESS as n grows, the result of the main theorem cannot simply be invoked directly.

For comparison purposes the gains in efficiency that result from using the generalized MESS instead of $S1_{\text{corr}}$ are carefully analyzed. In particular, for a security level of 2^{-s} , using MESS results in shorter stegotexts as long as the 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 the min-entropy of D is at least 17. The gains in encoding and decoding efficiency are even more dramatic. This is because $S1_{\text{corr}}$ needs expensive error-correcting codes, while MESS

simply repeatedly samples 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.

Stateless Multibit Extension. The third contribution of this work is to prove that for memoryless covertext distributions D , MESS can securely transmit multibit hidden-text messages using bit-by-bit steganographic encoding *without any 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. Prior to this work, no analysis of such a stateless multibit construction was available in the literature. However, Hopper previously conjectured [5] that the advantage of an adversary against the multibit version of $S1_{\text{corr}}$ grows quadratically in the number of hiddentext bits sent. The multibit bound presented in this work, while derived in an attempt to disprove this conjecture, in fact, proves it.

It should be stressed that multi-bit result presented here only applies to distributions D that are memoryless, i.e., where each successive covertext message drawn from D is independent of the history of previously sent covertexts. By a non-trivial extension of the techniques used to bound the flaw of $S1_{\text{orig}}$ and to prove the security of MESS, this work proves that for any hiddentext message of length l bits, when using n covertexts to encode each hiddentext bit, the distinguishing advantage of the adversary is at most $6l^2p^n$ (plus an amount that remains negligible for reasonable lengths l).

1.3 Subsequent Work

Subsequent joint research by this author and Reyzin has shown that the construction MESS, described herein, can also be shown secure for general channels having more complicated conditional distributions. The only requirement is that for all possible prior histories h of messages from the sender to the recipient, the sender know a constant min-entropy lower

bound for the channel distribution conditioned on h . This is equivalent to their knowing a constant upper bound p for the probability of the most likely element in channel distribution.

The proof techniques used in the subsequent work employ a hybrid argument and are consequently less algebraic, and more general than those presented herein. The same proof works equally well for any distribution D with maximal probability p . Consequently, the new work, titled “Simple, Stateless Steganography” [11], subsumes the contents of this work.

It remains to prove that the bound on security achieved by MESS is nearly optimal. Specifically, it would be desirable to demonstrate that when the only information available to a stego-encoder is a lower bound on the min-entropy of the channel distribution, no construction can perform significantly better than MESS. This is currently being investigated.

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

2.1 Definitions

The main definitions and notational conventions from [7] utilized herein are reiterated below. Many of these are taken nearly verbatim from the original work. Three significant differences are: i) this work treats stegosystem reliability as a construction parameter rather than a fixed value; ii) the stego-decoder is not given oracle access to the channel; and iii) the message history h will generally be suppressed and omitted.

A unidirectional *channel* \mathcal{C} is defined to be a distribution of bit sequences. \mathcal{C}_h denotes the distribution of such sequences conditioned on the history of previously drawn bits h . Similarly, \mathcal{C}_h^B denotes the distribution of sequences of B -bit blocks drawn from \mathcal{C} conditioned on history h . For notational convenience all channel messages are assumed to be of fixed length B bits. Furthermore, it is assumed there exists an oracle M which on input h efficiently samples B -bit blocks from \mathcal{C}_h^B . Where the specific history h is irrelevant, M will

be used for $M(h)$.

For brevity, notation will be slightly abused and $D \stackrel{\text{def}}{=} \mathcal{C}_h^B$ will be used in place of $M(h)$, particularly when denoting oracle access to the distribution or when the sampler is not of primary interest². However, throughout this work it is assumed that parties accessing D know only a lower bound for the min-entropy of D and whatever statistical inferences they make sampling D . The distribution consisting of n sequential samples drawn from D will be denoted D^n . With respect to the original channel notation, $D^n \stackrel{\text{def}}{=} \mathcal{C}_h^{nB}$ (recall that \mathcal{C}_h^{nB} denotes a conditional distribution of messages of fixed length nB bits conditioned on history h).

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 shared secret key $K \in \{0, 1\}^\kappa$, a string $m \in \{0, 1\}^*$ (the *hiddentext*), a sent message history h , and a channel sampling oracle $M(h)$ for the distribution D (the *coverttext* distribution);
 $SE^{M(\cdot)}(K, m, h)$ returns a sequence of blocks from D , $x_1 \| x_2 \| \dots \| x_l$ (the *stegotext*).
2. SD takes as input³ the shared secret key K and a sequence of blocks $x_1 \| x_2 \| \dots \| x_l$ (the *stegotext*) from D ;
 $SD(K, x)$ returns a hiddentext m .

In the following $S(\kappa)$ denotes an instantiation of S with security parameter κ .

²Technically $M(h)$ is made necessary by the fact that neither complete knowledge of nor direct access to the distribution \mathcal{C}_h^B is assumed in this work.

³In [7] the stego-decoder SD keeps a record of the channel history h and is given access to the channel sampling oracle $M(h)$. However, none of their constructions for passive wardens, in particular the one generalized herein, makes use of this. Thus, for the sake of simplicity, in this work it is not assumed that SD has such access.

Stegosystem Reliability. For a stegosystem S with security parameter κ sending l bits hiddentext on channel C , the l -bit *reliability* of S is defined as

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

In general more reliable stegosystems are more interesting, since they will likely be of greater practical use.

The Adversary. In this work only non-malicious (passive) adversaries, those which do not alter messages between the sender and recipient, mounting *chosen hiddentext attacks* against S are considered. The capabilities of such an adversary are described next.

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

1. runs in time t .
2. samples D (\mathcal{C}_h^B) via the sampling oracle $M(h)$ for any chosen history h .
3. makes at most q queries of combined length l bits to an oracle $\mathcal{O}_b(\cdot, \cdot)$, where $b \in \{0, 1\}$ is a randomly chosen bit. $\mathcal{O}_0(m, h)$ returns $x \leftarrow \mathcal{C}_h^{|SE(K,m,h)|}$, whereas $\mathcal{O}_1(m, h)$ returns $x \leftarrow SE(K, m, h)$.
4. outputs bit $\hat{b} \in \{0, 1\}$, a guess for the value of b .

W , sometimes referred to as the “warden”⁴, is nothing more than a distinguisher in the usual sense: it attempts to differentiate the covertext and stegotext distributions. In the following, $\mathcal{W}(t, q, l)$ denotes the set of all adversaries W making at most q queries of total length at most l bits and running in time at most t .

⁴The idea of the adversary as a warden and the use of W to designate it is a consequence of original problem formulation in [12].

Stegosystem Advantage and Insecurity. The distinguishing *advantage* of a passive adversary W conducting a chosen hiddentext attack against stegosystem S with security parameter κ for a given channel C is defined as

$$\mathbf{Adv}_{S(\kappa),C}^{\text{SS}}(W) = \left| \Pr_{K \leftarrow \{0,1\}^\kappa; r \leftarrow \{0,1\}^*} [W_r^{M, \mathcal{O}_1(\cdot, \cdot)} = 1] - \Pr_{r \leftarrow \{0,1\}^*} [W_r^{M, \mathcal{O}_0(\cdot, \cdot)} = 1] \right|.$$

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

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

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

The following additional notation will be used to describe the randomness utilized by the constructions discussed herein. Let $U(B)$ denote the uniform distribution on the set of B -bit strings, and $U(B, 1)$ denote the uniform distribution on functions from B -bit strings to a single bit. The constructions also make use of pseudorandomness. The related notation follows next.

Pseudorandom Functions For key $K \in \{0, 1\}^\kappa$ and $n \in \mathbb{N}$, let $F_{K,n}$ denote a specific member of the pseudorandom function family⁵

$$\mathcal{F}(\kappa, n) : \{0, 1\}^\kappa \times \{0, 1\}^{nB} \rightarrow \{0, 1\}$$

having key K , input length nB bits, and output length 1 bit. When $n = 1$, $\mathcal{F}(\kappa)$ and F_K will be used in place of $\mathcal{F}(\kappa, n)$ and $F_{K,n}$.

⁵Pseudorandom predicates and functions are defined, among other places, in [3], for example.

For a probabilistic adversary A and fixed value n , the *PRF-advantage* of A over \mathcal{F} is defined as

$$\mathbf{Adv}_{\mathcal{F}(\kappa,n)}^{\text{PRF}}(A) = \left| \Pr_{K \leftarrow U(\kappa), r \leftarrow \{0,1\}^*} [A_r^{F_{K,n}(\cdot)} = 1] - \Pr_{G \leftarrow U(nB,1), r \leftarrow \{0,1\}^*} [A_r^{G(\cdot)} = 1] \right|.$$

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

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

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

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

$$H_\infty(D) = \min_{x \in D} \left\{ -\log_2 \Pr_D[x] \right\}.$$

For conditional distributions D , the minimum should also be over all possible histories h . Correspondingly, let $p \stackrel{\text{def}}{=} H_\infty(D)$ denote an upper bound on the probability of the most likely element in D over all histories h of previously drawn elements.

2.2 Problematic Construction

The “flawed” version of Construction 1 from Section 3.2 of [7] is presented below, both in the interest of self-containment and in order to help concretely ground the discussion of the flaw in its security analysis. Additionally, as the analysis of this work will demonstrate, in many cases Construction 1 can be used *without any modification*, and even when modification is necessary, it will be very minimal.

For simplicity, the following presentation of Construction 1, henceforth referred to as $S1_{\text{orig}}$ for brevity, is for one-bit hiddentext messages only. However, as previously mentioned, the authors of [7] extend it to multibit messages simply by using a synchronized counter. Multibit messages will be discussed in greater detail in Section 5.

$S1_{\text{orig}}$ relies on a *rejection sampler* sub-procedure RS, presented of in Section 2.3 of [7]. This rejection sampler, and consequently $S1_{\text{orig}}$, requires sampling access to the covertext distribution D , which again is assumed to consist of B -bit long strings. This is denoted by oracle access to $M(\cdot)$, a sampling oracle for D or, when notation is abused to D itself. RS also requires access to a predicate F whose domain is the support of D , that is, $F : \{0, 1\}^B \rightarrow \{0, 1\}$.

On input a target bit b and maximum number of allowed iterations *count*, RS draws covertexts x from D until one is found that evaluates to the target bit b under F , or until the maximum number of allowed iterations is reached. In the latter case RS outputs the last message drawn. The parameter k should be thought of as the security parameter for RS. The reliability of $S1_{\text{orig}}$, not surprisingly, depends directly on k . The specification of RS follows.

Procedure $RS^{M(\cdot), F(\cdot)}(b, k)$:

$i = 0$

repeat:

$x \leftarrow M(\cdot); i \leftarrow i + 1$

until $F(x) = b$ or $k = i$

Output: x

The stego-encoding algorithm SE for $S1_{\text{orig}}$ takes a key K for a pseudorandom function F_K , additional security parameter k for RS, hiddentext bit m , and channel history h as input. While in [7], $k = |K| = \kappa$ is a single security parameter, in this work they are separated for clarity. $S1_{\text{orig}}$ runs RS with input m and $|K|$ and returns the sample output by RS. The stego-decoding or extraction algorithm SD for $S1_{\text{orig}}$ takes the shared secret key K and a stegotext x and outputs the image of x under F_K as the hiddentext m .

Procedure $S1_{\text{orig}}.SE(K, k, m, h)$:

$x \leftarrow \text{RS}^{M(h), F_K(\cdot)}(m, k)$

$h \leftarrow h \| x$

Output: x

Procedure $S1_{\text{orig}}.SD(K, x)$:

$m \leftarrow F_K(x)$

Output: m

From here on the message history h and sampling oracle $M(h)$ will no longer be explicitly mentioned when discussing RS, SE , and SD .

2.3 Problem With the Construction

Corollary 1 in [7] falsely states that $S1_{\text{orig}}$ is steganographically secure *on all channels* \mathcal{C} with minimum entropy $H_\infty(D = \mathcal{C}_h^B) > 2$ against wardens W asking 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 $S1_{\text{orig}}$ by the insecurity of the pseudorandom function family \mathcal{F} . The authors were made aware of this issue by [9], and later provided a modified version in [6] which is denoted $S1_{\text{corr}}$ from here on.

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 , 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 $H_\infty(D) > 2$, fixed bit b , fixed $k \in \mathbb{N}$, and randomly chosen predicate G from $U(B, 1)$, the distribution of messages $x \in D$ output by $\text{RS}^{D, G}(b, k)$ 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 $S1_{\text{orig}}$ adversary W has advantage equal to the advantage of pseudorandom function F_K adversary A . In the first step, the proof shows $\text{RS}^{D, F_K} \approx \text{RS}^{D, G}$,

then in the second step infers $\text{RS}^{D,G} = D$ using *False Claim 1*, and thus concludes the advantages are equal from the 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 $\text{RS}^{D,G}$ 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

Hopper, Langford, and von Ahn corrected the flaw of S1_{orig} shortly after its publication. In [6] they detail the corrected version S1_{corr} , also described below, and prove its security. No analysis as to the extent of the flaw was provided along with the revised construction. The main result of this work, presented in Section 3, shows that the magnitude of the adversary advantage against S1_{orig} can be correctly and precisely quantified. The second result, presented in Section 4, shows that by using an appropriate generalization of S1_{orig} this distinguishing advantage of the adversary can, in fact, be made negligible for *any* distribution D .

There are two main differences between S1_{corr} and S1_{orig} . First, although S1_{corr} uses the same rejection sampler RS as S1_{orig} , it forces RS to stop after only $k = 2$ unsuccessful attempts. In this case, the output distribution of RS can be shown, as in [6] or using Lemma 1 of this work, to be identical to the coartext distribution D . Unfortunately, as the authors point out, limiting RS to two attempts increases the probability Δ that an encoding error is introduced by $\text{RS}^{D,FK^{(\cdot)}}(b, 2)$ to $\Delta = \frac{1}{2} - \frac{1-p}{4}$ (plus the pseudorandom function insecurity). Recall that p is the maximal 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 $H_\infty(D) \geq 1$. Essentially, the stego-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 $S1_{\text{corr}}$. In order to achieve reliable (i.e. $\mathbf{Rel} \approx 1$) hiddentext transmission, prior to stego-encoding $S1_{\text{corr}}$ first encodes the hiddentext input using an error correcting code that corrects Δ fraction of errors. The stego-decoder $S1_{\text{corr}}.SD$, in turn, as its final step reconstructs the transmitted hiddentext from the error-correcting codewords recovered.

3 Bounding the Magnitude of the Problem

Despite the seemingly bad news that the rejections sampler perceptibly alters non-uniform covertext distributions D , this work bounds the magnitude of the distortion by providing an upper bound on the statistical difference between D and $RS^{D,G(\cdot)}$.

3.1 Insecurity Upper Bound

Before presenting the formal theorem statement, some additional additional notation is needed. For D a distribution on B -bit strings and a function $G : D \rightarrow \{0, 1\}$, define α_G to be the weight of G where

$$\alpha_G \stackrel{\text{def}}{=} \sum_{x \in D: G(x)=1} \Pr_D[x],$$

and $\beta_G \stackrel{\text{def}}{=} 1 - \alpha_G$ to be the weight of the complement. Similarly, for a subset $S \subseteq D$, define $\alpha_S \stackrel{\text{def}}{=} \sum_{x \in S} \Pr_D[x]$ and $\beta_S \stackrel{\text{def}}{=} 1 - \alpha_S$. Lastly, define

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

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

Theorem 1. *Let D be any discrete probability distribution with highest probability p , $k \in \mathbb{N}$, and $b \in \{0, 1\}$ a bit. For a randomly chosen predicate $G : D \rightarrow \{0, 1\}$, the statistical difference between D and $\text{RS}^{D,G}(b, k)$ is at most $2p$ plus a negligible function in k . More precisely,*

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

3.2 Supporting Results

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, consider the limiting case when the maximum number covertext samples drawn by RS, the parameter k in the above, is allowed to go to infinity. Note that in S1_{orig} , the cutoff parameter k for RS is taken to be length of the pseudorandom function key K , i.e. $k = \kappa = |K|$. However, from this point forward, k and κ will be considered independent parameters.

The following lemma provides an expression for the probability distribution of $\text{RS}^{D,G}$ for infinite k . Lemma 2 then uses this expression to give an analogue to Theorem 1 for infinite k .

Lemma 1. *For x an element from the support of D and bit $b \in \{0, 1\}$, define $\text{RS}^{D,G}(b, \infty) \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \text{RS}^{D,G}(b, k)$, and*

$$\Pr_{G \in U(B,1), D}[\text{RS}^{D,G}(b, \infty) \rightarrow x] \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \Pr_{G \in U(B,1), D}[\text{RS}^{D,G}(b, k) \rightarrow x].$$

Then,

$$\Pr_{G \in U(B,1), D}[\text{RS}^{D,G}(b, \infty) \rightarrow x] = \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{G \in U(B,1): G(x)=1} \frac{1}{\alpha_G} \right)$$

where the probability is taken over the choice of G .

Proof. The proof of this Lemma is contained in Appendix A. \square

The next step is the formulation of the infinite analog of Theorem 1 which is used later in its proof.

Lemma 2. *Let D be any discrete probability distribution with highest probability p and $b \in \{0, 1\}$ a bit. For a randomly chosen predicate $G : D \rightarrow \{0, 1\}$, the statistical difference between D and $\text{RS}^{D,G}(b, \infty)$ is at most $2p$. More precisely,*

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

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, \dots, a_n ,*

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

Proof. The proposition can be verified by recalling that the *harmonic mean* of a set of n values a_1, a_2, \dots, a_n , is defined as $n/(1/a_1 + \dots + 1/a_n)$, whereas the usual *arithmetic mean* is defined as $(a_1 + \dots + a_n)/n$. 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. \square

Proof of Lemma 2. First, recall 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, let $t \stackrel{\text{def}}{=} |D|$.

For each function G , 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,

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

Thus,

$$\Pr_D[x] - \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(b, \infty) \rightarrow x] \leq \Pr[x] - \frac{\Pr[x]}{1 + \Pr[x]} = \frac{(\Pr[x])^2}{1 + \Pr[x]} \leq (\Pr[x])^2.$$

Finally, combining these two pieces,

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

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

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

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

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

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

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

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

4 MESS: A Secure Generalization of the Problematic Construction

Theorem 1 of the preceding section shows that for D with sufficiently high min-entropy, S1_{orig} (i.e., Construction 1 of [7]) needs no modification. On the other hand, since p is fixed for any given D , the error of S1_{orig} is not a negligible function. Thus, when D lacks sufficiently high min-entropy, S1_{orig} in its current form is *insecure*. This is the motivation for the second contribution of this work: MESS, a generalized version of S1_{orig} that is secure *for all* D . MESS stands for “Minimum-Entropy-Sensitive Stegosystem.”

4.1 Construction Description

The problem with S1_{orig} is that it can only utilize whatever min-entropy D provides. To remedy this limitation, this work modifies RS to make use of the well known technique of

repeated sampling on D in order to effectively increase the minimum entropy of D . This modified version of the rejection sampler is denoted RS-HE. Specifically, instead of using one covert message $x \in D$ per hiddentext bit, RS-HE uses n sequentially drawn covert messages $x_1, x_2, \dots, x_n \in D$. The ordered concatenation of these n covert messages is then evaluated under the predicate F (with a suitably expanded domain). The exact value of n depends on $H_\infty(D)$, or equivalently on the maximal probability p , and is fixed for a given D . The proposed stegosystem MESS is the same as $S1_{\text{orig}}$ except for a few minor syntactic changes necessary to accommodate the 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 shared, secret pseudorandom predicate key K , the maximum number of sampling rounds attempted by RS-HE, and the number of covert message samples drawn per round from D and used to encode each hiddentext bit. Let $\text{MESS}(\kappa, k, n)$ denote the new system instantiated with these parameters. A formal description of MESS can be found in Appendix C.

4.2 MESS Security

The security proof for $S1_{\text{orig}}$ given in [7] only attempts to show security with respect to adversaries making a single 1-bit query. This section will do the same for MESS. Recall that multi-bit security can follow from 1-bit security by adding a synchronized counter as in [7]. However, by adapting the proof techniques developed in this section for the 1-bit case, Theorem 4 in Section 5 proves something stronger. Namely, it proves that MESS can securely send multibit messages statelessly, that is, *without a synchronized counter*, in the special case where the covert message distribution D is memoryless.

The proof that MESS is 1-bit steganographically secure follows, although not immediately, from Theorem 1 with D^n in place of D . Recall that D^n denotes the distribution

formed by successively drawing n sequential elements from D , conditioned, if necessary, on the history of previously drawn elements h ; draw i is conditioned on h and the $i - 1$ draws which precede it. Clearly, the first term in Theorem 1 becomes at most p^n and can be made negligible by taking n sufficiently large. Recall that p is the maximal probability in D over all possible histories h . The only complication is that the second term, $\eta(D^n, k) = 2^{-|D^n|} \sum_{S \subseteq D^n} \alpha_S^k$ now depends on both n and k . It remains to show that this second term can still be made negligible, even as n grows.

Theorem 2. *Let D be a discrete probability distribution conditioned on history h of previously drawn elements, and let p be the probability of the most likely element of D ($p = 2^{-H_\infty(D)}$) taken over all possible histories h . For any $0 < \delta < 1/2$,*

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

Before proving Theorem 2, the issue of bounding the term $\eta(D^n, k)$ which appears in the proof, will be dealt with. It turns out 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. $\zeta(D^n, k)$ can be bound in two steps. First, 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 . Then, Lemma 5 bounds ζ of this uniform distribution. As will be seen in Lemma 6 (in Section 4.3), $\zeta(D^n, k)$ is exactly the failure probability of the rejection sampler RS-HE $^{D, G}$.

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 $H_\infty(D)$, define*

$U_D = U(\lfloor 2^{H_\infty(D)} \rfloor)$, that is U_D is a uniform distribution with $\lfloor 2^{H_\infty(D)} \rfloor$ elements. Then for all $k \in \mathbb{N}$, $\zeta(D, k) \leq \zeta(U_D, k)$.

Proof. The proof of this Lemma is contained in Appendix D. \square

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 $0 < \delta < \frac{1}{2}$,

$$\zeta(U(t), k) \leq \left(\frac{1}{2} + \delta\right)^k + e^{-2t\delta^2}.$$

Proof. The proof of this Lemma is contained in Appendix D. \square

Proof (sketch) of Theorem 2. As previously indicated, the proof follows by applying the result of Theorem 1 to the distribution D^n , using the combined results of Lemma 4 and Lemma 5 to bound the resulting $\eta(D^n, k)$ term, and accounting for the advantage due to a pseudorandom function F . A more detailed version of the proof can be found in Appendix D. \square

4.3 MESS Reliability

An explicit bound on the insecurity **InSec** of the stegosystem MESS was provided in the previous section. However, another important stegosystem property to consider is the reliability **Rel**, that is, the probability that the recipient correctly decodes the encoded message. The stegosystem definition in [7] formally requires only **Rel** $\geq 2/3$, however, in reality communicating parties will most likely desire **Rel** ≈ 1 . The following theorem bounds the reliability of MESS.

Theorem 3. Let D be a discrete probability distribution conditioned on history h of previously drawn elements, and let p be the probability of the most likely element of D ($p =$

$2^{-H_\infty(D)}$) taken over all possible histories h . For any $0 < \delta < 1/2$,

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

The following lemma bounds the probability of encoding error in the simpler case of $\text{RS}^{D, G}$ (equivalently $\text{RS-HE}^{D, G}$ with $n = 1$) for a random function G .

Lemma 6. *Let D be any discrete probability distribution and $b \in \{0, 1\}$ a bit. For a randomly chosen predicate $G \leftarrow U(|D|, 1)$, the encoding error introduced by $\text{RS}^{D, G}(b, k)$ 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. $\text{RS}^{D, G}(b, k)$ introduces encoding error whenever after k unsuccessful attempts to find a coartext $x \in D$ such that $G(x) = b$, it outputs the last (k th) sample x drawn from D . Using algebra similar to that in the proof of Lemma 1, this probability can be shown to be $\zeta(D, k)$. \square

Proof of Theorem 3. The reliability of $\text{MESS}(\kappa, k, n)$ is simply one minus the encoding error introduced by $\text{RS-HE}^{D, F_{K, n}}(\cdot, k, n)$ where $F_{K, n}$ is a pseudorandom predicate with key K on the domain D^n . Recall that in the proof of Theorem 2 it was argued that $\text{RS-HE}^{D, F_{K, n}}(\cdot, k, n)$ and $\text{RS}^{D^n, F_K}(\cdot, k)$ 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 $\text{RS-HE}^{D, F_{K, n}}(\cdot, k, n)$ is at most $\zeta(D^n, k) + \mathbf{InSec}_{\mathcal{F}(\kappa, n)}^{\text{PRF}}(O(nk), k)$. 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 pseudorandom function. Using the upper bound for $\zeta(D^n, k)$ from (16) in the proof of Theorem 2 and subtracting from one gives the indicated lower bound for the reliability. \square

4.4 MESS Parameter Choices and Efficiency

Given covertext distribution D with min-entropy $H_\infty(D)$, 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 $n = \lceil (s + 2)/H_\infty(D) \rceil$, $k = s + 6$, and κ such that for the chosen pseudorandom function family $\mathcal{F}(\kappa, n)$, $\mathbf{InSec}_{\mathcal{F}(\kappa, n)}(O(nk), k) \leq 2^{-s-3}$. A precise derivation of these parameter values can be found in Appendix E. The length of the stegotext output by MESS is just n covertexts long.

In comparison, as Appendix F shows, to achieve reasonable reliability, $\mathbf{S1}_{\text{corr}}$ needs to send more than 5 covertexts per hiddentext bit. For distributions with very low min-entropy even more covertexts are needed per bit. Thus, if $H_\infty(D) \geq (s + 2)/5$, MESS sends fewer covertexts than $\mathbf{S1}_{\text{corr}}$, and if $H_\infty(D) \geq (s + 2)$, MESS sends only a single covertext, thereby effectively reducing to $\mathbf{S1}_{\text{orig}}$. Moreover, MESS requires no computationally expensive error-correction.

5 Stateless, Multibit Extension of MESS for Memoryless Distributions

As was previously mentioned, a secure stateful multibit version of MESS can readily 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 pseudorandom function $F_{K,n}$. The counter essentially serves to refresh the pseudorandom function key, thereby making each successive hiddentext bit as secure as the first. However, as the next theorem shows, if the covertext message distribution D is memoryless, direct bit-by-bit encoding with MESS yields a secure and stateless stego-encoding of multibit hiddentext messages. This eliminates the need for a synchronized counter, provided D is memoryless. Here memoryless means that each draw from D is

completely independent of any history h of previous draws.

Theorem 4. *Let D be a memoryless covertext message distribution, and let p be the probability of the most likely element of D ($p = 2^{-H_\infty(D)}$). For $l \geq 1$ total hiddentext bits queried by the adversary W ,*

$$\mathbf{InSec}_{\text{MESS}(\kappa, k, n), D}^{\text{SS}}(t, l, l) \leq 6l^2 p^n + 4l \left(\left(\frac{1}{2} + \delta \right)^k + e^{-\lfloor \frac{t}{p^n} \rfloor 2\delta^2} \right) + \mathbf{InSec}_{\mathcal{F}(\kappa, n)}^{\text{PRF}}(t + O(\ln k), lk).$$

As was done in the proof of Theorem 2, intermediate results for the simpler case of $n = 1$ (effectively S1_{orig}) will be developed first and then generalized for MESS with $n > 1$. In the following let $\text{RS}^{D, G}(1^l 0^l, \infty) \rightarrow x$ denote the event that the rejection sampler on bit-by-bit input $1^l 0^l$, outputs $x = x_1 x_2 \dots x_{2l}$, where each $x_i \in D$ and $G : D \rightarrow \{0, 1\}$ is a randomly chosen predicate.

Proof (sketch) of Theorem 4. 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 l hiddentext bits can be bound by the advantage of a *non-adaptive* adversary asking $2l$ -bit hiddentext queries of the form $1^l 0^l$. The second, Lemma 8, shows that a sequence of elements $x = x_1 x_2 \dots x_{2l}$ containing no repeated elements is no less likely to occur as a stego-encoding of $1^l 0^l$ than as a random draw from D^{2l} . This is contingent on the sampler $\text{RS}^{D, G}$ being allowed to make as many draws as needed, i.e. $k = \infty$.

From there, using two corollaries of Lemma 8, Lemma 9 bounds the statistical difference between the infinite version of $\text{RS}^{D, G}$ and D^{2l} , by considering the relative probability of sequences of $2l$ messages with and without collisions drawn from D versus those output by $\text{RS}^{D, G}$. Finally, Lemma 10 deals with the remaining statistical difference between the

infinite and finite versions of the rejection sampler, and then the advantage associated with the pseudorandom function is accounted for. \square

5.1 Supporting Results for Multibit Insecurity Upper Bound

The statement of the aforementioned lemmas and corollaries follows. Their proofs can be found in Appendix G. Following these supporting lemmas is a more detailed proof of Theorem 4.

Lemma 7. *Let D be any memoryless discrete probability distribution and W' an adversary that adaptively asks the stego-encoding oracle for $MESS(\kappa, k, n)$ operating on D for the bit-by-bit stego-encoding of a total of l bits of hiddentext. The advantage of this l -bit adaptive adversary W' is bound by the advantage of a non-adaptive adversary W that asks the same oracle for the bit-by-bit stego-encoding of the $2l$ -bit hiddentext $1^l 0^l$.*

Proof. The proof of this Lemma is contained in Appendix G. \square

Lemma 8. *For all $x = x_1 x_2 \cdots x_{2l} \in D^{2l}$ such that $\forall i, j \ x_i \neq x_j$, that is for all strings of $2l$ elements from a memoryless distribution D which contains no repeated element,*

$$\Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \geq \Pr_{D^{2l}}[x].$$

Proof. The proof of this Lemma is contained in Appendix G. \square

The following two corollaries follow directly from analysis of collision probabilities in the two distributions and from Lemma 8.

Corollary 1 (To Lemma 8: Non-collision Statistical Difference). *For a memoryless distribution D the statistical difference between $\text{RS}^{D,G}(1^l 0^l, \infty)$ and D^{2l} for elements $x = x_1 x_2 \cdots x_{2l} \in D^{2l}$ such that no value x_i is repeated, i.e. $\forall i, j$ where $1 \leq i \neq j \leq l \ x_i \neq x_j$,*

is less than probability of drawing an element x from D^{2l} containing at least one repeated element. Namely,

$$\sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j x_i \neq x_j} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| \leq \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j x_i = x_j} \Pr_{D^{2l}}[x].$$

Proof. The proof of this Corollary is contained in Appendix G. \square

Corollary 2 (To Lemma 8: Collision Statistical Difference). *For a memoryless distribution D the probability that $\text{RS}^{D,G}(1^l 0^l, \infty)$ outputs an element $x = x_1x_2 \cdots x_{2l}$ of D^{2l} , such that at least one value x_i is repeated, i.e. $\exists i, j$ such that $1 \leq i \neq j \leq l$ and $x_i = x_j$, is less than or equal to the probability of drawing such an x from D^{2l} directly.*

Proof. The proof of this Corollary is contained in Appendix G. \square

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 $1^l 0^l$ for any $1 \leq l$ and a randomly chosen predicate $G : D \rightarrow \{0, 1\}$, the statistical difference between D^{2l} and $\text{RS}^{D,G}(1^l 0^l, \infty)$ is at most $6l^2 p$. More precisely,*

$$\sum_{\forall x \in D^{2l}} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| \leq 6l^2 p.$$

Proof. The proof of this Lemma is contained in Appendix G. \square

Lemma 10. *Let D be any memoryless discrete probability distribution. For a fixed $k, l \in \mathbb{N}$,*

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

Proof. The proof of this Lemma is contained in Appendix G. \square

Proof of Theorem 4. The structure of the proof is similar to that of Theorem 2. The proof follows by first inserting positive and negative $\Pr_{G \in U(B,1),D}[\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x]$ 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 2. Finally, adjusting for the advantage due to a pseudorandom F gives the desired result. \square

5.2 Optimality of Multibit Bound

The bound of Theorem 4 is nearly optimal, as the following theorem shows. Basically, an adversary asking $l/2$ 1-queries followed by $l/2$ 0-queries can distinguish with probability roughly $l^2 p^n / 4$.

Theorem 5. *For any $0 < p < 1$, there exists a discrete probability distribution D for which $\max_{x \in D} \{\Pr_D[x]\} = p$ and such that for a randomly chosen predicate $G : D \rightarrow \{0, 1\}$, the statistical difference between D^l and $\text{RS}^{D,G}(1^{l/2} 0^{l/2}, \infty)$ is greater than one half the probability of obtaining a collision among l draws from D . That is,*

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

The proof which follows 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 $\text{RS}^{D,G}$ and non-zero for D^l .

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 $x_1 \dots x_l$ 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 $\text{RS}^{D,G}(1^{l/2} 0^{l/2}, \infty)$.

Now in the case of D^l , 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, lower bounding the probability of collisions among an l element draw from D in general, can be accomplished by upper bounding the probability of non-collisions as follows,

$$\sum_{x=x_1x_2\cdots x_l \in D^l \mid \forall i \neq j \ x_i \neq x_j} \Pr_{D^l}[x] = (1-p)(1-2p)\cdots(1-(l-1)p) \quad (1)$$

$$\leq e^{-p-2p-\cdots-(l-1)p} \quad (2)$$

$$= e^{-pl(l-1)/2} \quad (3)$$

$$\leq 1 - pl(l-1)/2 + (pl(l-1)/2)^2/2. \quad (4)$$

Line (2) and Line (4) follow from the Taylor series expansion of e^{-x} which gives $(1-x) \leq e^{-x} \leq 1-x+x^2/2$. Thus, the probability of collisions among the l elements drawn from D is,

$$\begin{aligned} \sum_{x=x_1x_2\cdots x_l \in D^l \mid \exists 1 \leq i \leq l/2 < j \leq l \ x_i = x_j} \Pr_{D^{2l}}[x] &= 1 - \sum_{x=x_1x_2\cdots x_l \in D^l \mid \forall i \neq j \ x_i \neq x_j} \Pr_{D^{2l}}[x] \\ &\geq 1 - \left(1 - \frac{pl(l-1)}{2} + \frac{(pl(l-1)/2)^2}{2} \right) \\ &= \frac{pl(l-1)}{2} - \frac{(pl(l-1)/2)^2}{2}. \end{aligned}$$

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

A Proof of Lemma 1

Proof. First the case of $b = 1$ will be proven and then it will be argued by symmetry that this also suffices to prove the case of $b = 0$. To compute the probability that $\text{RS}^{D,G}(1, k)$ outputs x , one need simply find the expected value over the $2^{|D|}$ possible random functions $G : D \rightarrow \{0, 1\}$, as follows,

$$\begin{aligned} \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1, k) \rightarrow x] &= \frac{1}{2^{|D|}} \left(\sum_{G:G(x)=1} \Pr_D[x] \sum_{i=0}^{k-1} \beta_G^i + \sum_{G:G(x)=0} \Pr_D[x] \beta_G^{k-1} \right) \\ &= \frac{\Pr_D[x]}{2^{|D|}} \left(\sum_{G:G(x)=1} \frac{1 - \beta_G^k}{1 - \beta_G} + \sum_{G:G(x)=0} \beta_G^{k-1} \right). \end{aligned} \quad (5)$$

Taking the limit as $k \rightarrow \infty$, that is as the rejection sampler makes greater and greater numbers of draws from D before “giving up”, gives

$$\begin{aligned} \lim_{k \rightarrow \infty} \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1, k) \rightarrow x] &= \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{G:G(x)=1} \frac{1}{1 - \beta_G} \right) \\ &= \frac{\Pr_D[x]}{2^{|D|}} \left(1 + \sum_{G:G(x)=1} \frac{1}{\alpha_G} \right). \end{aligned}$$

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 ,

$$\Pr[\text{RS}^{D,G}(0, k) \rightarrow x] = \Pr[\text{RS}^{D,\hat{G}}(1, k) \rightarrow x].$$

Generalizing this over all possible choices for the function G gives

$$\Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(0, k) \rightarrow x] = \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1, k) \rightarrow x].$$

So, considering $\text{RS}^{D,G}(1, k)$ is sufficient, and the proof is complete. \square

Remark 1. It can be seen from (5) and some algebra, that when $k = 2$, in fact,

$\Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(b, k) \rightarrow x] = \Pr_D[x]$ 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 (5) from the proof of Lemma 1 it follows that

$$\sum_{\forall x \in D} \left| \Pr_{G \in U(B,1),D}[\text{RS}^{D,G}(b, \infty) \rightarrow x] - \Pr_{G \in U(B,1),D}[\text{RS}^{D,G}(b, k) \rightarrow x] \right| \quad (6)$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| 1 + \sum_{S \subseteq D: x \in S} \frac{1}{\alpha_S} - \sum_{S \subseteq D: x \in S} \frac{\beta_S^k - 1}{\beta_S - 1} - \sum_{S \subseteq D: x \notin S} \beta_S^{k-1} \right| \quad (7)$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| 1 + \sum_{S \subseteq D: x \in S} \frac{1}{\alpha_S} - \sum_{S \subseteq D: x \in S} \frac{1 - \beta_S^k}{\alpha_S} - \sum_{S \subseteq D: x \in S} \alpha_S^{k-1} \right| \quad (8)$$

$$= \sum_{\forall x \in D} \frac{\Pr[x]}{2^{|D|}} \left| \sum_{S \subsetneq D: x \in S} \frac{\beta_S^k - \alpha_S^k}{\alpha_S} \right| \quad (9)$$

$$= 2 \sum_{x \in D: |\cdot| \geq 0} \frac{\Pr[x]}{2^{|D|}} \sum_{S \subsetneq D: x \in S} \frac{\beta_S^k - \alpha_S^k}{\alpha_S} \quad (10)$$

$$\leq \frac{1}{2^{|D|-1}} \sum_{\forall x \in D} \Pr[x] \sum_{S \subsetneq D: x \in S} \frac{\beta_S^k}{\alpha_S} \quad (11)$$

$$= \frac{1}{2^{|D|-1}} \sum_{S \subsetneq D: S \neq \emptyset} \frac{\beta_S^k}{\alpha_S} \sum_{\forall x \in S} \Pr[x] \quad (12)$$

$$= \frac{1}{2^{|D|-1}} \sum_{S \neq \emptyset} \beta_S^k = \frac{1}{2^{|D|-1}} \sum_{S \subsetneq D} \alpha_S^k \quad (13)$$

$$\leq 2\eta(D, k). \quad (14)$$

Line (8) follows from the definitions of α and β and the symmetry of the set of all functions. To obtain Line (9), combine the sums and remove the term 1 by restricting S to be a *proper* subset of D . Line (10) follows from the same property of statistical difference used in the proof of Lemma 2. Line (12) 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 the last line which completes the proof. \square

C Formal Description of MESS

C.1 For Memoryless Distributions

For now, assume that the covertext distribution D 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 suppressed.

Let n be an additional security parameter for MESS and RS-HE. It specifies the number elements of D (coverttexts) over which a single hiddentext bit will be encoded. Recall that $S1_{\text{orig}}$ and RS had security parameters $\kappa = |K|$ and k , the length of the pseudorandom predicate key and the maximum number of sampling attempts made by RS, respectively. Like RS, RS-HE uses a predicate F , however the domain is expanded, i.e. now $F : D^n \rightarrow \{0, 1\}$. When running as a subroutine of MESS, RS-HE has oracle access to $F_{K,n}$, a member of the pseudorandom predicate family $\mathcal{F}(\kappa, n) : \{0, 1\}^\kappa \times \{0, 1\}^{nB} \rightarrow \{0, 1\}$ with key K and input length nB bits. Recall that each element in D is assumed to be B bits long. The modified version of RS is:

Procedure RS-HE ^{$D, F(\cdot)$} (b, k, n):

```

i = 0
repeat :
  for j = 1 to n:
     $x_j \leftarrow D$ 
   $x \leftarrow (x_1|x_2|\dots|x_n)$ 
  i  $\leftarrow i + 1$ 
until  $F_{K,n}(x) = b$  or i = k

```

Output: x

The only differences between the stego-encoding algorithms for MESS and for $S1_{\text{orig}}$ is that $\text{MESS}.SE$ has additional input n that is used when calling RS-HE, and its stegotext output is n times longer. The stego-decoding algorithm $\text{MESS}.SD$ is unchanged from $S1_{\text{orig}}.SD$ except that its stegotext input is n times longer. It should be emphasized that with respect to $S1_{\text{orig}}$ given in Section 2.2, if history is irrelevant, 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 use of $F_{K,n}$ instead of F_K , and the n times longer stegotext output by SE and input to SD .

C.2 For General Distributions

To generalize the modifications of the previous section, it is necessary to drop the assumption that D is a memoryless distribution. Suppose instead that the distribution of covertexts is *conditional* on the history h of previously sent messages. The distribution resulting from sending n messages is more complex than $D \times D \times \cdots \times D$. Although it is an abuse of notation, 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 .

Formal Description of 1-bit MESS

Procedure RS-HE ^{$M(h), F_{K,n}(\cdot)$} (b, k, n):

$i = 0$

repeat :

$h' \leftarrow h$

for $j = 1$ to n :

$$y_j \leftarrow M(h')$$

$$h' \leftarrow h'|y_j$$

$$x \leftarrow (y_1|y_2|\dots|y_n)$$

$$i \leftarrow i + 1$$

until $F_{K,n}(x) = b$ **or** $j = k$

$$h \leftarrow h'$$

Output: x

Procedure $\text{MESS.SE}(K, n, k, m, h)$:

$$x \leftarrow \text{RS-HE}^{M(h), F_{K,n}(\cdot)}(m, k, n)$$

$$h \leftarrow h|x$$

Output: x

Procedure $\text{MESS.SD}(K, n, x)$:

$$m \leftarrow F_{K,n}(x)$$

Output: m

Remark 2. The inner “for” loop of RS-HE can be thought of as an oracle $M^n(h)$ —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 $M(h)$. Thus, the analysis of RS given in Theorem 1 applies here as well, except that D must be replaced with D^n .

D Proofs of Lemma 4 and 5, and Theorem 2

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. \square

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 yields

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

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

Proof of Lemma 4. D can be transformed into U_D by adding mass to the highest-probability elements of D until their probability reaches $1/\lfloor 2^{H_\infty(D)} \rfloor$, while simultaneously removing the same mass from the lowest-probability elements until their probability reaches 0. By Claim 2, ζ of the resulting distribution will not decrease. Then, removing all zero-probability elements gives U_D which, by Claim 1, will not change ζ . \square

Proof of Lemma 5. Consider ζ as a subset of the 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 none of those gets selected in the k tries. More precisely, rewriting the definition of ζ ,

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

The exponential term follows from the application of Hoeffding's Inequality⁶ [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,

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

□

Proof of Theorem 2. First the case of MESS for a truly random predicate F is considered, and then, the necessary correction for a pseudorandom F will be added. The security of MESS is completely determined by the security of RS-HE and the pseudorandom random predicate $F_{K,n}$ which it accesses.

Recall that D^n is the coartext distribution consisting of n subsequent draws from the given coartext distribution D via its sampling oracle $M(h)$, where the input h is the history of previously drawn elements from D . Let $M^n(h)$ denote an efficient sampling oracle for D^n . As was pointed out in Remark 2 at the end of Section C.2, such an M^n can easily be

⁶The use of such a bound makes sense since for $S \subseteq U(t)$, $t\alpha_S = |S|$, that is the number of heads/ones observed for on t independent fair coin tosses.

constructed from M and, in fact, $\text{RS-HE}^{M(\cdot), F(\cdot)}(b, k)$ is equivalent to $\text{RS}^{M^n(\cdot), F(\cdot)}(b, k)$ for the same predicate F . Thus, applying Theorem 1 gives,

$$\begin{aligned} & \sum_{\forall x \in D^n} \left| \Pr_{D^n}[x] - \Pr_{F \in U(nB, 1), M}[\text{RS-HE}^{M(\cdot), F(\cdot)}(b, k, n) \rightarrow x] \right| \\ &= \sum_{\forall x \in D^n} \left| \Pr_{D^n}[x] - \Pr_{F \in U(nB, 1), M}[\text{RS}^{M^n(\cdot), F(\cdot)}(b, k) \rightarrow x] \right| \\ &\leq 2p^n + 2\eta(D^n, k) \end{aligned} \tag{15}$$

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

Clearly the first term in (15) 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 $t = \lfloor p^{-n} \rfloor$ gives

$$\begin{aligned} \eta(D^n, k) &< \zeta(D^n, k) \\ &\leq \left(\frac{1}{2} + \delta \right)^k + e^{-\lfloor p^{-n} \rfloor 2\delta^2}. \end{aligned} \tag{16}$$

Finally, combining (15) and (16) and accounting for the advantage due to the pseudorandom function $F_{K, n}$,

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

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

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

□

E MESS Parameter Derivation and Running Time

By Theorem 2, the insecurity of MESS when sending a single hiddentext bit is at most

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

It remains to derive useful values of the parameters κ , k , and n for MESS. Suppose the goal is for MESS to operate on a given coverttext distribution D having min-entropy $H_\infty(D) > 1$, or equivalently maximal probability $p = 2^{-H_\infty(D)}$, with at most 2^{-s} insecurity and a corresponding reliability of at least $1 - 2^{-s}$. First, take $n \geq (s + 2)/H_\infty(D)$, which ensures that $2p^n < 2^{-s-1}$, and $k = s + 6$. Then, for $\delta = 1/(4(s + 4))$, 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} , it is necessary for $\lfloor 1/p^n \rfloor 2\delta^2 \log_2 e \geq s + 4$. Substituting 2^{s+2} for $1/p^n$ and $1/(4(s + 4))$ for δ , gives that $2^{s+2} \log_2 e \geq 8(s + 4)^3$ is needed, which holds as long as $s \geq 13$. Note that insecurity greater than 2^{-13} is generally not acceptable in most applications, so this is not a serious restriction.

Finally, κ is chosen so that the insecurity $\mathbf{InSec}_{\mathcal{F}(\kappa, n)}^{\text{PRF}}(t + O(nk), k)$ of the given pseudorandom function family $\mathcal{F}(\kappa, n)$ 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 sampling rounds RS-HE attempts, but the *expected* number of rounds is just 2.

For each hiddentext bit, the stego-encoder for MESS essentially just draws, on average, $2n$ samples from the coverttext distribution D and thus evaluates, on average, twice the pseudorandom predicate $F_{K, n}$ on the concatenation of n samples. Similarly, for each hiddentext bit, the stego-decoder of MESS just evaluates $F_{K, n}$ on the stegotext received, i.e., on the concatenation of the n messages from D . Thus, the running time of its decoder is essentially one pseudorandom function evaluation, and the average running time of its encoder is about twice that. The stegotext length is clearly just n coverttexts long.

Final Values: In summary, 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 $n \geq \lceil (s+2)/H_\infty(D) \rceil$, $k = s+6$, and κ such that for the chosen pseudorandom function family \mathcal{F} , $\text{InSec}_{\mathcal{F}(\kappa,n)}^{\text{PRF}}(t+O(nk), k) \leq 2^{-s-3}$. Observe that to achieve 2^{-80} security and $1 - 2^{-80}$ reliability, if $H_\infty(D) \geq 82$ then *MESS has a stegotext only one coverttext long*, that is, MESS simplifies to S1_{orig} .

F Revised Construction 1 Parameter Derivation

Here it will be argued that for secure and reliable transmission, S1_{corr} needs to send $1 - H_2(1/2 - 1/4(1-p))$ coverttexts per hiddentext bit, where H_2 is the binary entropy. This value is between 5 and 6 for reasonable p , specifically those which are less than 0.05.

The error correcting codes needed by S1_{corr} to assure reliable hiddentext transmission⁷ will stretch each hiddentext bit to be sent 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

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

G Proofs of Lemmas 7 and 8, Corollaries 1 and 2, and Lemmas 9 and 10

Proof of Lemma 7. Suppose W' adaptively asks for the bit-by-bit encoding of the message $m = m_1 m_2 \dots m_l$, $m_i \in \{0, 1\}$. First, W simply asks its corresponding stego-encoding oracle for the bit by bit encoding of $1^l 0^l$. Then, W uses the encoding of the i th zero or the

⁷It bears reiterating 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, a useful system will generally require much higher reliability. Therefore, for comparison purposes, this work requires that stegosystems be reliable with probability close to 1.

i th one that it received to answer the i th adaptively chosen query of W' . 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. \square

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, \dots, 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 (5) from the proof of Lemma 1 gives,

$$\Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x_1 x_2 \cdots x_{2l} \text{ s.t. } \forall i, j \ x_i \neq x_j] \quad (17)$$

$$= \frac{\Pr_{D^{2l}}[x]}{2^{|D|}} \sum_{S \subset D: x_1, x_2, \dots, x_l \in S \wedge x_{l+1}, \dots, x_{2l} \notin S} \frac{1}{\alpha_S^l \beta_S^l} \quad (18)$$

$$= \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \frac{1}{2^{|D|-2l}} \sum_S \frac{1}{\alpha_S^l (1 - \alpha_S)^l} \quad (19)$$

$$\geq \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \left(\frac{\sum_S \frac{1}{\alpha_S (1 - \alpha_S)}}{2^{|D|-2l}} \right)^l \quad (20)$$

$$\geq \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \left(\frac{2^{|D|-2l}}{\sum_S \alpha_S (1 - \alpha_S)} \right)^l \quad (21)$$

$$\geq \frac{\Pr_{D^{2l}}[x]}{2^{2l}} \left(\frac{2^{|D|-2l}}{\sum_{i=1}^{2^{|D|-2l}} 1/4} \right)^l \quad (22)$$

$$= \frac{\Pr_{D^{2l}}[x] 4^l}{2^{2l}} \quad (23)$$

$$= \Pr_{D^{2l}}[x]. \quad (24)$$

Line (20) follows from Proposition 2, line (21) from Proposition 1, and line (22) from the fact that $\max_{0 < \alpha < 1} \alpha(1 - \alpha) = 1/4$, and the remaining lines follow from algebra. \square

Proof of Corollary 1. 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,

$$\begin{aligned}
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \left| \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| = \\
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \left(\Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right) = \\
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \\
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \Pr_{D^{2l}}[x] \\
& = \left(1 - \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \right) - \\
& \left(1 - \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{D^{2l}}[x] \right) \\
& = \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{D^{2l}}[x] - \\
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \\
& \leq \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{D^{2l}}[x].
\end{aligned}$$

□

Proof of Corollary 2. Since by Lemma 8, for every string x of $2l$ unique elements from D ,

$$\Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \geq \Pr_{D^{2l}}[x],$$

$$\begin{aligned}
& \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] = \\
& 1 - \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \Pr_{G \in U(B,1),D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \\
& \leq 1 - \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \forall i,j \ x_i \neq x_j} \Pr_{D^{2l}}[x] \\
& = \sum_{x=x_1x_2\cdots x_{2l} \in D^{2l} \mid \exists i,j \ x_i = x_j} \Pr_{D^{2l}}[x].
\end{aligned}$$

□

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,

$$\begin{aligned}
& \sum_{\forall x \in D^{2l}} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| = \\
& \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \forall i,j \mid x_i \neq x_j} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| + \\
& \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \exists i,j \mid x_i = x_j} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] - \Pr_{D^{2l}}[x] \right| \\
& \leq \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \exists i,j \mid x_i = x_j} \Pr_{D^{2l}}[x] + \\
& \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \exists i,j \mid x_i = x_j} \left| \Pr_{G \in U(B,1), D} [\text{RS}^{D,G}(1^l 0^l, \infty) \rightarrow x] \right| + \\
& \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \exists i,j \mid x_i = x_j} \left| \Pr_{D^{2l}}[x] \right| \\
& \leq 3 \sum_{x=x_1 x_2 \dots x_{2l} \in D^{2l} \mid \exists i,j \mid x_i = x_j} \Pr_{D^{2l}}[x] \\
& \leq 6l^2 p.
\end{aligned}$$

□

Proof of Lemma 10. This sum captures the difference in probabilities between the rejection sampler in the infinite and finite cases. The element $x = x_1 x_2 \dots x_{2l}$ 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. □

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 International 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 Association for Computing Machinery*, 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, 2003.
- [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*, volume 2442 of *Lecture Notes in Computer Science*. Springer-Verlag, 18–22 August 2002. Corrected version 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] Lea Kissner, Tal Malkin, and Omer Reingold. Private communication to N. Hopper, J. Langford, L. von Ahn, 2002.
- [10] Tri Van Le and Kaoru Kurosawa. Efficient public key steganography secure against adaptively chosen stegotext attacks. Technical Report 2003/244, Cryptology e-print archive, <http://eprint.iacr.org>, 2003.
- [11] Leonid Reyzin and Scott Russell. Simple, stateless steganography. Technical Report 2003/093, Cryptology e-print archive, <http://eprint.iacr.org>, 2004.
- [12] 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.
- [13] Luis von Ahn and Nicholas J. Hopper. Public-key steganography. Technical Report 2003/233, Cryptology e-print archive, <http://eprint.iacr.org>, 2003. To appear in *Advances in Cryptology—EUROCRYPT 2004*.