On Steganographic Chosen Covertext Security

Nicholas Hopper

University of Minnesota
4-192 EECS, 200 Union St SE, Minneapolis MN 55455
hopper@cs.umn.edu

Abstract. At TCC 2005, Backes and Cachin proposed a new and very strong notion of security for public key steganography: secrecy against adaptive chosen covertext attack (SS-CCA); and posed the question of whether SS-CCA security was achievable for *any* covertext channel. We resolve this question in the affirmative: SS-CCA security is possible for any channel that admits a secure stegosystem against the standard and weaker "chosen hiddentext attack" in the standard model of computation. Our construction requires a public-key encryption scheme with ciphertexts that remain indistinguishable from random bits under adaptive chosen-ciphertext attack. We show that a scheme with this property can be constructed under the Decisional Diffie-Hellman assumption. This encryption scheme, which modifies a scheme proposed by Kurosawa and Desmedt, also resolves an open question posed by von Ahn and Hopper at Eurocrypt 2004.

1 Introduction

Suppose that Alice and Bob are prisoners, and that their prison warden has foolishly allowed them to send "harmless messages" between their cells, so long as he may listen to everything they say. *Steganography* is the study of techniques that allow Alice and Bob to hide arbitrary messages – *hiddentexts* – in their apparently harmless communications (normally, *covertexts*) so that the warden cannot detect the presence of these messages. The case where the prisoners share a secret key has been studied extensively in both information-theoretically [5] and computationally secure settings [13, 9]. Several recent papers have also addressed the case in which one or both of the prisoners has a public key [1, 3, 17]. *In this paper, we are only concerned with the* **bare public key scenario***, considered in [3], in which only Bob publishes a public key, and any prisoner can send hidden information to Bob.*

A recent paper by Backes and Cachin [3] considers the scenario where the warden may also inject messages into the channel between Alice and Bob, and observe Bob's reaction to these messages. Roughly, [3] gives a formal model of this scenario and defines a strong sense of security against this adversary: a stegosystem is said to be *steganographically secure against adaptive chosen covertext attacks* (SS-CCA) if, even in this case, the warden cannot tell whether Alice's messages contain hiddentexts. Analogously to the standard cryptographic notion of a chosen ciphertext attack, this seems to be the most general type of attack possible on a system for steganography.

Backes and Cachin leave open the problem of constructing a stegosystem satisfying SS-CCA, and instead address a relaxed notion of security, against adaptive *replayable* chosen-covertext attacks (SS-PDR-CCA). Roughly, in this notion, the warden is still allowed to inject messages into the channel between Alice and Bob, *except* that he is now restricted from sending messages which are, in some sense, *replays* of previous messages sent by Alice. Intuitively, two covertexts are replays of each other with respect to a public key if they decode to the same hiddentext. Backes and Cachin construct public-key stegosystems which satisfy SS-PDR-CCA under a variety of assumptions.

While it is an important advancement to limit the adversary to replay attacks, these attacks still constitute a serious threat against steganography. Imagine that Alice sends Bob some message which prompts an "unusual" reaction; in a replay attack, the warden can construct an apparently harmless covertext which corresponds to the same hiddentext as Alice's message, and send it to Bob. If Bob has the same "unusual" reaction, in response to a *different message*, it suggests to the warden that Alice's covertext contained a hidden message.

In this paper, we show how the previously known schemes fail in defending against replay attacks, and modify them to demonstrate the feasibility of the SS-CCA security condition, for any *efficiently sampleable* channel. This is a stronger assumption on the channel than in many previous works on steganography [1, 18, 9, 3], which assume only oracle access to the channel distribution. However, [14] shows that any channel which admits a secure stegosystem at all (in the standard model of computation) must be efficiently sampleable. Thus this construction serves as a demonstration that the SS-CCA notion is *feasible*, even though our particular construction may not always be practical to implement.

Our construction relies on the existence of public-key encryption schemes which are pseudorandom against chosen-ciphertext attack, a nonstandard security notion for encryption schemes. We also show that such encryption schemes exist, *without need of the random oracle assumption*,[1] under the Decisional Diffie-Hellman assumption. The existence of an encryption scheme satisfying this notion was an open question posed by von Ahn and Hopper [1].

Related Work. In addition to the work of Backes and Cachin [3], which we build on, Le and Kurosawa [17] and von Ahn and Hopper [1] have both proposed notions of security against "chosen stegotext attack." The notion proposed in [17] seems to be equivalent to SS-CCA; however the construction proposed there requires that the receiver know the sender's public key in order to decode. Similarly, the SS-CSA notion of [1] explicitly includes the public key of the sender; it can be thought of as an "attacker-specific" notion of security. However, the security model of [1] is also intended to prevent forgery by the warden, which is not a concern in the present model.

Both of these schemes require the sender to publish a public key. While this may not be a concern for ordinary communication, it is undesirable for steganography. This is because the aim of the sender in steganography is to *avoid suspicion* – yet publishing a public key for a stegosystem may be inherently suspicious.

[1] We note that several constructions in the random oracle model are known [4, 19].

On the other hand, it is frequently the case, as [1] argue, that the receiver of steganography need not avoid suspicion. This could be the case when, for example, the receiver is a newspaper or government agency wishing to receive whistle-blowing reports. Or when the receiver is a human-rights organization that would like to receive reports from its volunteers in the field. Thus it is important to have a construction which is secure in the bare public key model.

Other recent papers on foundations of steganography have focused on the private key setting. Cachin [5] formulated a model for steganography in an information-theoretic setting. Hopper *et al* [13] gave the first rigorous formulation of steganography with computational security, and demonstrated the feasibility of the notion with provably secure constructions. They also proposed the model of communication which subsequent work has followed. Independently, Katzenbeisser and Petitcolas [15] proposed a similar security condition. Dedić *et al* [9] address bounds on communication rate for a generic stegosystem. Lysyanskaya and Meyerovich [18] consider the possibility of an imperfect covertext oracle.

Anderson and Petitcolas [2] first proposed the possibility of public-key steganography and gave a heuristic construction. Craver [8] proposed a notion of public-key steganography with heuristic security against removal of the hiddentext. von Ahn and Hopper [1] were the first to formulate rigorous security definitions for the public-key case and demonstrate that public-key steganography was feasible.

Notation A function $\mu : \mathbb{N} \to [0, 1]$ is said to be *negligible* if for every $c > 0$, for all sufficiently large n, $\mu(n) < 1/n^c$. We denote the length (in bits) of a string or integer s by $|s|$. The concatenation of string s_1 and string s_2 will be denoted by $s_1 \| s_2$. The assignment $a \|_l b = c$ means that a is the first l bits of c and b is the remaining $|c| - l$ bits of c. We assume the existence of efficient, unambiguous *pairing* and *un-pairing* operations, so (s_1, s_2) is not the same as $s_1 \| s_2$.

We let U_k denote the uniform distribution on k bit strings. If V denotes an event in some probability space, we denote its complement by $\overline{\mathsf{V}}$. If $\mathcal{D}$ is a probability distribution with finite support X, we define the *minimum entropy* of $\mathcal{D}$, by $H_\infty(\mathcal{D}) = \min_{x \in X}\{\log_2(1/\Pr_{\mathcal{D}}[x])\}$. For a probability distribution $\mathcal{D}$, we denote by $x \leftarrow \mathcal{D}$ the action of drawing a sample x according to $\mathcal{D}$. We denote the statistical difference between distributions $\mathcal{D}$ and $\mathcal{E}$, with finite support X, by $\|\mathcal{D} - \mathcal{E}\| = \frac{1}{2}\sum_{x \in X} |\Pr_{\mathcal{D}}[x] - \Pr_{\mathcal{E}}[x]|$.

2 Pseudorandomness against chosen-ciphertext attack

We will need to construct a public-key encryption scheme which satisfies a non-standard security notion: indistinguishability from random bits under chosen-ciphertext attack. A scheme satisfying this notion is also non-malleable [10] and has pseudoranom ciphertexts [1]; the existence of a scheme simultaneously satisfying these latter notions *without random oracles* was an open question posed by von Ahn and Hopper at Eurocrypt 2004 [1].

Let $\mathcal{E}$ be a public-key encryption scheme with message expansion function ℓ. We define a chosen-ciphertext attack against $\mathcal{E}$ as a game played by an oracle adversary A:

1. $A^{D_{SK}}(PK)$ outputs *challenge message* $m^* \in \{0,1\}^{l^*}$.
2. A is given a *challenge ciphertext* c^*, where either $c \leftarrow E_{PK}(m^*)$ or $c \leftarrow U_{\ell(l^*)}$.
3. A continues to query D_{SK} subject to the restriction that A may not query $D_{SK}(c^*)$. A outputs a bit.

We define A's CCA advantage against $\mathcal{E}$ by

$$\mathbf{Adv}^{\mathsf{cca}}_{\mathcal{E},A}(k) = \left|\Pr[A^{D_{SK}}(PK, E_{PK}(m^*)) = 1] - \Pr[A^{D_{SK}}(PK, U_\ell) = 1]\right| ,$$

where $m^* \leftarrow A^{D_{SK}}(PK)$ and $(PK, SK) \leftarrow G(1^k)$, and define the CCA insecurity of $\mathcal{E}$ by $\mathbf{InSec}^{\mathsf{cca}}_{\mathcal{E}}(t,q,\mu,l^*,k) = \max_{A\in\mathcal{A}(t,q,,\mu,l^*)}\left\{\mathbf{Adv}^{\mathsf{cca}}_{\mathcal{E},A}(k)\right\}$, where $\mathcal{A}(t,q,\mu,l^*)$ denotes the set of adversaries running in time t, that make q queries of total length μ, and issue a challenge message m^* of length l^*. Then $\mathcal{E}$ is (t,q,μ,l^*,k,ϵ)-*indistinguishable from random bits under chosen ciphertext attack* if $\mathbf{InSec}^{\mathsf{cca}}_{\mathcal{E}}(t,q,\mu,l^*,k) \leq \epsilon$. $\mathcal{E}$ is called *indistinguishable from random bits under chosen ciphertext attack* (IND$-CCA) if for every probabilistic polynomial time (PPT) A, $\mathbf{Adv}^{\mathsf{cca}}_{A,\mathcal{E}}(k)$ is negligible in k.

We show a simple modification of an encryption scheme of Kurosawa and Desmedt [16] (which itself is a modification of the original Cramer-Shoup encryption scheme [7]) which satisfies IND$-CCA. The main modification to the scheme is to use a dense encoding of the DDH subgroup and rejection sampling to produce uniform k-bit strings.

Setup. We let p_k, Q_k be large primes such that $p = 2Q+1$ and $2^{k+1} > Q > 2^k$. We let $g \in \mathbb{Z}_p^*$ have order Q, and define the maps $lr : \langle g\rangle \to \mathbb{Z}_Q$, $qr : \mathbb{Z}_Q \to \langle g\rangle$ such that $lr(v) = v$ if $v \leq Q$ and $lr(v) = -v \bmod p$ otherwise; and $qr(u) = u$ if u is a quadratic residue modulo p and $qr(u) = p-u$ otherwise. Notice that $qr \circ lr$ is the identity map on the quadratic residues and $lr \circ qr$ is the identity map on $\mathbb{Z}_Q$. We assume the Decisional Diffie Hellman (DDH) assumption: for any PPT A, $\mathbf{Adv}^{\mathsf{ddh}}_{A,g,p,Q}(k) = |\Pr_{x,y\leftarrow\mathbb{Z}_Q}[A(g^x, g^y, g^{xy}) = 1] - \Pr_{x,y,z\leftarrow\mathbb{Z}_Q}[A(g^x, g^y, g^z) = 1]|$ is negligible.

We assume the existence of a family of target collision-resistant hash functions $H : \{0,1\}^{2k} \to \mathbb{Z}_Q$,[2] A universal family of hash functions $\Lambda : \mathbb{Z}_Q \to \{0,1\}^{2k'}$, an IND$-CPA symmetric-key encryption scheme E, D with k'-bit keys,[3] and a pseudorandom function family $F : \{0,1\}^{k'} \times \{0,1\}^* \to \{0,1\}^\tau$.[4] Note that the existence of all of these primitives is implied by the DDH assumption.

Key Generation. Choose random $g_1, g_2 \in \langle g\rangle$, and choose random $x_1,x_2,y_1,y_2 \in \mathbb{Z}_Q$. Compute the group elements $c = g_1^{x_1} g_2^{x_2}, d = g_1^{y_1} g_2^{y_2}$. Choose hash functions H, Λ. The public key is $(g_1, g_2, c, d, H, \Lambda)$ and the private key is (x_1, x_2, y_1, y_2).

[2] so for any PPT A, $\mathbf{Adv}^{\mathsf{tcr}}_{A,H}(k) = \Pr_{h\leftarrow H}[h(A(h(x))) = x : x \leftarrow \mathbb{Z}_Q]$ is negligible

[3] so for any PPT A, $\mathbf{Adv}^{\mathsf{cpa}}_{A,E}(k) = |\Pr_{K\leftarrow U_{k'}}[A^{E_K}(1^{k'}) = 1] - \Pr[A^{U_{|\cdot|}}(1^{k'}) = 1]|$ is negligible

[4] so for any PPT A, $\mathbf{Adv}^{\mathsf{prf}}_{A,F}(k) = |\Pr_{K\leftarrow U_{k'}}[A^{F_K}(1^{k'}) = 1] - \Pr_{f:\{0,1\}^*\to\{0,1\}^\tau}[A^f(1^{k'}) = 1]|$ is negligible.

Encryption. Given a message $m \in \{0,1\}^*$, repeat the following steps:

- Choose $r \leftarrow \mathbb{Z}_q$.
- Compute $u_1 = lr(g_1^r)$, $u_2 = lr(g_2^r)$

Until u_1, u_2 are both at most 2^k. Then compute $\alpha = H(u_1 \| u_2)$, $v = c^r d^{r\alpha}$, $(K, \kappa) = \Lambda(v)$, $e = E_K(m)$, $T = F_\kappa(e)$. The ciphertext is $u_1 \| u_2 \| e \| T$.

Decryption. To decrypt the ciphertext $u_1 \| u_2 \| e \| T$, first compute $\alpha = H(u_1 \| u_2)$ and compute $v = qr(u_1)^{x_1 + y_1 \alpha} qr(u_2)^{x_2 + y_2 \alpha}$, $K \| \kappa = \Lambda(v)$. Test whether $T = F_\kappa(e)$; if not output $\perp$, otherwise output $D_K(e)$.

Theorem 1. *If $k \geq 4k'$, then*

$$\mathbf{InSec}^{\mathsf{cca}}_{\mathcal{E}}(t,q,\mu,l,k) \leq 8\mathbf{InSec}^{\mathsf{tcr}}_{H}(t,k) + 12\mathbf{InSec}^{\mathsf{ddh}}_{g,p,Q}(t) + 4\mathbf{InSec}^{\mathsf{cpa}}_{E}(t,1,l,k') + (16q+4)\mathbf{InSec}^{\mathsf{prf}}_{F}(t,q,\mu,k') + 8q(2^{-\tau} + 2^{-k'+1}) + 2^{-k'+4}$$

The security proof appears in the full version and closely follows the security proof for Kurosawa and Desmedt's scheme given by Gennaro and Shoup [11].

3 Definitions

Channels. We follow previous work [13, 17, 1, 9] in modeling the communication between two parties by a *channel.* We define a channel $\mathcal{C}$ as a family of probability distributions on documents from a set D, indexed by sequences $h \in D^*$. A channel implicitly defines an indexed distribution on sequences of ℓ documents — given index h, draw $d_1 \leftarrow \mathcal{C}_h$, $d_2 \leftarrow \mathcal{C}_{(h,d_1)}$, ..., $d_\ell \leftarrow \mathcal{C}_{(h,d_1,\ldots,d_{\ell-1})}$. We call the index h the *history* and label this distribution on sequences by $\mathcal{C}_h^\ell$. A history $h = (d_1, d_2, \ldots, d_\ell)$ is called *legal* (denoted $h \in \mathcal{H}$) if for all i, $\Pr_{\mathcal{C}_{(d_1,\ldots,d_{i-1})}}[d_i] > 0$. A channel is *always informative* if for every legal history h, $H_\infty(\mathcal{C}_h^\ell) = \Omega(\ell)$.

We will require that a channel be *efficiently sampleable:* there is an efficiently computable algorithm $\mathsf{channel}$ such that $\mathsf{channel}(h, U_k)$ and $\mathcal{C}_h$ are computationally indistinguishable.[5] This is in contrast to the models of [13, 9, 1, 3], where the channel is assumed to be accessible only via a probabilistic oracle. While results in that model are in some sense more general, we refer the reader to [14] for a proof that in the standard model of computation, sampleability is necessary for secure steganography.

Since it is widely believed that all natural processes can be computed in probabilistic polynomial time [12], we do not *in theory* rule out steganography for any realistic channels by requiring the channel to be sampleable. On the other hand, it is conceivable that there are channels which we can currently sample physically but not computationally, and thus in practice it is still an open problem to design a stegosystem which is SS-CCA secure for such channels.

[5] Some examples of widely used channels satisfying this notion include: scientific simulations, cryptography and security protocols, computer games, financial modeling, weather forecasts, etc.

Public-Key Stegosystem. A public-key stegosystem $\mathcal{S}$ is a triple of probabilistic algorithms:

- $\mathcal{S}.\mathsf{Generate}$ (abbreviated SG) takes as input a security parameter 1^k and generates a key pair $(\rho, \sigma) \in \mathcal{PK} \times \mathcal{SK}$.
- $\mathcal{S}.\mathsf{Encode}$ (abbreviated SE) takes as input a public key $\rho \in \mathcal{PK}$, a string $m \in \{0,1\}^*$ (the *hiddentext*), and a channel history h. $SE(\rho, m, h)$ returns a sequence of documents $s_1, s_2, \ldots, s_l$ (the *stegotext*) from the support of $\mathcal{C}_h^l$.
- $\mathcal{S}.\mathsf{Decode}$ (abbreviated SD) takes as input a secret key $\sigma \in \mathcal{SK}$, a sequence of documents $s_1, s_2, \ldots, s_l$, and a channel history h. $SD(\sigma, s, h)$ returns a hiddentext $m \in \{0,1\}^*$.

We require that a stegosystem is *correct:* for every polynomial $p(k)$ there exists a negligible $\nu(k)$ such that for every $m \in \{0,1\}^{p(k)}$, legal history h, and $(\rho, \sigma) \in [SG(1^k)]$, $\Pr[SD(\sigma, SE(\rho, m, h), h) = m] \geq 1 - \nu(k)$.

Chosen-Covertext Attack. In an adaptive chosen-covertext attack against a public-key stegosystem $\mathcal{S}$, a challenger draws a key pair $(\rho, \sigma) \leftarrow SG(1^k)$, and an adversary W is given PK and allowed oracle access to SD_σ. The attacker produces a *challenge hiddentext* m^* and history h^* and is given as a response a sequence of documents $s^* \in D^{\ell(|m^*|)}$. After this, the attacker continues to query SD with the restriction that he may not query $SD(s^*)$. (As always, W may depend on the channel distribution $\mathcal{C}$) At the conclusion of the attack, W must guess whether $s^* \leftarrow SE(\rho, m^*, h^*)$ or $s^* \leftarrow \mathcal{C}_{h^*}^{\ell^*}$. We define the (steganographic) *Chosen-Covertext Advantage* of W against $\mathcal{S}$ with respect to $\mathcal{C}$ by

$$\mathbf{Adv}^{\mathsf{scca}}_{\mathcal{S},\mathcal{C},\mathcal{W}}(k) = \left| \Pr[W^{SD_\sigma}(PK, SE(\rho, m^*, h^*)) = 1] - \Pr[W^{SD_\sigma}(\rho, \mathcal{C}_{h^*}^{l^*}) = 1] \right| \ ,$$

where $(m^*, h^*) \leftarrow W^{SD_\sigma}(\rho)$ and $(\rho, \sigma) \leftarrow SG(1^k)$. We define the sCCA insecurity of $\mathcal{S}$ with respect to $\mathcal{C}$ by

$$\mathbf{InSec}^{\mathsf{scca}}_{\mathcal{S},\mathcal{C}}(t, q, \mu, l^*, k) = \max_{W \in \mathcal{W}(t,q,\mu,l^*)} \left\{ \mathbf{Adv}^{\mathsf{scca}}_{\mathcal{S},\mathcal{C},\mathcal{W}}(k) \right\} \ ,$$

where $\mathcal{W}(t, q, \mu, l^*)$ denotes the class of all W running in time t which make at most q oracle queries of μ bits and submit a challenge hiddentext of length at most l^*.

We say that $\mathcal{S}$ is $(t, q, \mu, l, k, \epsilon)$ secure against chosen-covertext attack with respect to $\mathcal{C}$ if $\mathbf{InSec}^{\mathsf{scca}}_{\mathcal{S},\mathcal{C}}(t, q, \mu, l, k) \leq \epsilon$, and that $\mathcal{S}$ is secure against chosen-covertext attack with respect to $\mathcal{C}$ ($\mathsf{SS\text{-}CCA}$) if $\mathbf{Adv}^{\mathsf{scca}}_{\mathcal{S},\mathcal{C},\mathcal{W}}(k)$ is negligible for all PPT W.

4 Previous constructions

Both previously known constructions of (bare) public-key steganography [1, 3] have a common structure. Let $\mathcal{F}$ denote a strongly universal family of hash functions $f : D \to \{0,1\}$. Let $f \leftarrow \mathcal{F}$ be chosen as part of a public key, or fixed as a "common reference string." Then both constructions use the routine shown in figure 1 to hide uniformly chosen bits in $\mathcal{C}$

Procedure sample:
Input: target $c \in \{0,1\}$, history h, bound k
Let $j = 0$
repeat:
  sample $s \leftarrow \mathcal{C}_h$, increment j
until $f(s) = c$ OR $(j > k)$
Output: s

Fig. 1. Sampling routine

Proposition 1. *Let $\mathcal{C}$ be always informative and $f \leftarrow \mathcal{F}$. Then for any $h \in \mathcal{H}$,*

$$\|(f, \mathsf{sample}(h, U_1, k)) - (f, \mathcal{C}_h)\| \leq 2^{-H_\infty(C_h)/2} \ .$$

The proposition is a direct consequence of the leftover hash lemma. If the channel is always-informative, sample can operate on samples from $\mathcal{C}_h^k$ and induce only a negligible statistical difference in its output distribution. The basic construction of a stegosystem, HashRS, is shown in figure 2, where (G, E, D) is a public-key cryptosystem which has pseudorandom ciphertexts.

Informally, the scheme works by transforming the hiddentext into a uniform-looking ciphertext $c = E_{PK}(m)$. The ciphertext bits are then used one at a time (or w at a time, with sampling costs and statistical difference increased by a factor of 2^w) to select covertexts that hash (via f) to the bits of the ciphertext, using sample. Since the ciphertext looks uniform, the covertexts thus selected will be indistinguishable from samples from $\mathcal{C}_h$. Decoding applies f to each covertext document to recover the ciphertext c, and then decrypts this ciphertext using SK to compute the hiddentext $m = D_{SK}(c)$.

The Backes-Cachin construction instantiates HashRS with a public-key encryption scheme which satisfies two properties. First, it must be PDR-CCA secure, as defined by Canetti *et al* [6]. Second, the encryption scheme should have *pseudorandom* ciphertexts: given the public key it was encrypted under, a ciphertext should be computationally indistinguishable from a random string of the same length. When instantiated with a public-key cryptosystem satisfying these properties, we call the resulting stegosystem $\mathcal{BC}$.

Intuitively, the SS-PDR-CCA security of the $\mathcal{BC}$ scheme arises from the fact that W is disallowed from submitting covertexts that decode to the same hiddentext. Thus an attack W against $\mathcal{BC}$ can easily be turned into a PDR-CCA attack A against the underlying encryption scheme. The main technical step is in simu-

Procedure Encode:
Input: $m \in \{0,1\}^l$, h, PK
Draw $c_1 \cdots c_\ell \leftarrow E(PK, m)$
for $i = 1 \ldots \ell$ do
  set $s_i = \mathsf{sample}(c_i, (h, s_{1,\ldots,i-1}), k)$.
Output: $s_1, s_2, \ldots, s_\ell$

Procedure Decode:
Input: $s_1, s_2, \ldots, s_l$, SK
for $i = 1 \ldots l$ do
  set $c_i = f(s_i)$
set $c = c_1 || c_2 || \cdots || c_l$.
Output: $D(K, c)$

Fig. 2. HashRS Stegosystem

lating decryption queries: whenever W queries the decoding oracle on a covertext $s = s_1, \ldots, s_\ell$, the PDR-CCA attacker computes a ciphertext $c = c_1, \ldots, c_\ell$ by setting $c_i = f(s_i)$. If the ciphertext c is a replay of the challenge ciphertext c^*, then the stegotext s is also a replay, so A responds to W with $\perp$. Otherwise A queries his decryption oracle at c and returns the result to W.

This standard simulation technique also hints at a CCA attack against the $\mathcal{BC}$ stegosystem. We now formally describe the attack W. On input PK, W uniformly picks a challenge message $m^* \leftarrow U_{l^*}$. On receiving the challenge covertext s^*, W computes c^* by setting $c_i^* = f(s_i^*)$. W computes a "replay" covertext $s' \leftarrow \mathsf{sample}((h^*, s^*), c^*, k)$. Finally, W queries the decryption oracle on s'. If $SD_{SK}(s') = m^*$, W outputs 1 and otherwise W outputs 0. It is obvious that when $s^* \in SE(PK, m^*, h^*)$, then we will have that $SD(SK, s', h^*) = m^*$ except when encoding fails, since otherwise unique decryption requires that $D_{SK}(c^*) = D_{SK}(E_{PK}(m^*)) = m^*$. On the other hand, when $s^* \leftarrow \mathcal{C}_h^\ell$, then m^* and s^* are chosen independently of each other, so $\Pr[D_{SK}(c^*) = m^*] \leq 2^{-l^*}$.

Proposition 2. *For every l^*, there exists a negligible function $\nu(k)$ such that*

$$\mathbf{Adv}^{\mathsf{scca}}_{W,\mathcal{BC}}(k) \geq 1 - 2^{-l^*} - \nu(k)$$

Note that the "replay" covertext will be indistinguishable from a sample from the channel, so the decoder would have no reason not to decode it and act on any information contained in the hiddentext. Thus this attack is reasonable, in that it could be applied in a realistic scenario, rather than being merely an artifact of the model. Of course the adversary might further attempt to replay the exact stegotext; this latter attack is, however, impossible to defeat.

5 Our Construction

Intuitively, the reason the attack in the previous section succeeds is that even though the underlying *ciphertext* is non-malleable, there are many possible encodings of the ciphertext. This observation immediately suggests a possible improvement: design a sampling method such that each ciphertext corresponds to exactly one stegotext. Indeed, the construction of [17] seems to have this property, but this construction inherently requires a shared secret between the encoder and the decoder. Likewise, the "attacker-specific" construction of [1] seems to achieve a similar property, but validity of a stegotext is determined by the sender's public key. Our construction modifies this latter approach to remove this dependence on the sender, and also removes the reliance on the random oracle model from that construction.

We make use of the fact that we have an efficiently sampleable channel $\mathcal{C}$, and will make use of the "deterministic encoding" routine shown in figure 3. This algorithm works in a similar manner to the HashRS.Encode algorithm, with the exception that the randomness for sampling is an explicit argument. Thus for a given sequence of lk random inputs, this routine has exactly one possible encoding for any message $c \in \{0,1\}^l$. Thus if an l-bit, non-malleable, ciphertext can determine the lk bits of sampling randomness to be used in its encoding, we

Procedure DEncode:
Input: bits $c_1, \ldots, c_l$, history h, bound k, randomness $r_1, \ldots r_{lk} \in \{0,1\}^k$
Let $\iota = 0$; for $i = 1 \ldots l$ do
  Let $j = 0$; repeat:
    compute $s_i = \mathsf{channel}((h, s_{1\ldots i-1}), r_\iota)$; increment j, ι
  until $f(s_i) = c_i$ OR $(j > k)$
Output: $s_1, s_2, \ldots, s_l$

Fig. 3. Deterministic Encode

Procedure Encode:	**Procedure** Decode:
Input: $m \in \{0,1\}^\ell$, h, PK	**Input:** $s_1, \ldots, s_l$, h, SK
Choose $r \leftarrow U_k$	Let $c = f(s_1) \| \cdots \| f(s_l)$
Let $c = E_{PK}(r \| m)$	Let $r \|_k m = D_{SK}(c)$.
Let $\boldsymbol{r} = G(r)$	Set $\boldsymbol{r} = G(r)$.
Output: $\mathsf{DEncode}(c, h, k, \boldsymbol{r})$	If $s \neq \mathsf{DEncode}(c, h, k, \boldsymbol{r})$ return $\perp$.
	Output: m

Fig. 4. SCCA Stegosystem

can prevent replay attacks. One way to do this is to apply a random oracle to the randomness used in producing the ciphertext; this approach was used by [1]. We instead use a pseudorandom generator to expand an k-bit seed into an lk-bit sequence and then include this seed in the plaintext. Proving the security of this approach requires some additional care, because now it is conceivable that the sampling algorithm could leak information about the plaintext.

We now formally describe our construction. We will assume that $\mathcal{E}$ is a public-key IND$-CCA secure encryption scheme, and $(PK, SK) \leftarrow \mathcal{E}.G(1^k)$. Furthermore, we assume that for any l, $\Pr[D_{SK}(U_l) \neq \perp] \leq \nu(k)$ for some negligible ν. Thus, *valid* ciphertexts, which do not decrypt to $\perp$, have negligible density. For convenience, we assume that for all m, $|E_{PK}(m)| = \ell(|m|)$, for some polynomial ℓ. We will also assume that $G : \{0,1\}^k \to \{0,1\}^{k \times lk}$ is a pseudorandom generator. The final scheme SCCA is shown in Figure 4.

Theorem 2. *Let $f \leftarrow \mathcal{F}$ and let $\epsilon = \max_{h \in \mathcal{H}} \left\{ 2^{-H_\infty(\mathcal{C}_h^k)/2} \right\} = 2^{-\Omega(k)}$. Then*

$$\mathbf{InSec}^{\mathsf{scca}}_{\mathsf{SCCA},\mathcal{C}}(t, q, \mu, l, k) \leq \mathbf{InSec}^{\mathsf{cca}}_{\mathcal{E}}(t', q, \mu, l, k) + \nu(k) + \ell(l{+}k)\epsilon + \mathbf{InSec}^{\mathsf{prg}}_{G}(t', k),$$

where $t' \leq t + O(lk)$.

Proof. Choose an arbitrary $W \in \mathcal{W}(t, q, \mu, l)$; let $(PK, SK) \leftarrow G(1^k)$ and let $(m^*, h^*) \leftarrow W^{SD_{SK}}(PK)$. We will bound $\mathbf{Adv}^{\mathsf{scca}}_{W,\mathsf{SCCA},\mathcal{C}}(k)$ by considering the following sequence of hybrid distributions:

- D_1: $\mathcal{C}_{h^*}^{\ell(l+k)}$
- D_2: $\mathsf{DEncode}(U_{\ell(l+k)}, h^*, k, U_{k \times lk})$
- D_3: $\mathsf{DEncode}(U_{\ell(l+k)}, h^*, k, G(U_k))$
- D_4: $\mathsf{DEncode}(E_{PK}(r \| m^*), h^*, k, G(r))$, where $r \leftarrow U_k$

Clearly D_4 perfectly simulates the stegotext distribution, and likewise D_1 perfectly simulates the covertext distribution. For convenience, we will define the

quantity $\mathbf{Adv}^i_W(k) = \left|\Pr[W^{SD}(PK, D_{i+1}) = 1] - \Pr[W^{SD}(PK, D_i) = 1]\right|$, and note that

$$\begin{aligned}\mathbf{Adv}^{\mathsf{scca}}_{W,\mathsf{SCCA},\mathcal{C}}(k) &= \left|\Pr[W^{SD}(PK, D_4) = 1] - \Pr[W^{SD}(PK, D_1) = 1]\right| \\ &\leq \mathbf{Adv}^1_W(k) + \mathbf{Adv}^2_W(k) + \mathbf{Adv}^3_W(k).\end{aligned}$$

Thus we proceed to bound $\mathbf{Adv}^i_W(k)$ for $i \in \{1, 2, 3\}$.

Lemma 1. $\mathbf{Adv}^1_W(k) \leq \ell(l + k)\epsilon$

Proof. This follows because $\|f(C_h) - U_1\| \leq \epsilon$, and no (nonuniform) efficient process can increase statistical distance.

Lemma 2. $\mathbf{Adv}^2_W(k) \leq \mathbf{InSec}^{\mathsf{prg}}_G(t', k)$

Proof. We will construct a PRG adversary A for G such that $\mathbf{Adv}^{\mathsf{prg}}_{A,G}(k) = \mathbf{Adv}^2_W(k)$. A works as follows: first, A picks a key pair $(PK, SK) \leftarrow G(1^k)$ to use in responding to the queries W makes to SD. A is given as input a string $r \in \{0,1\}^{k \times lk}$ and asked to decide whether $r \leftarrow U_{k \times lk}$ or $r \leftarrow G(U_k)$. Then A can achieve advantage precisely $\mathbf{Adv}^2_W(k)$ by emulating W, responding to its decoding queries using SK, and responding to the challenge hiddentext (m^*, h^*) by drawing $c \leftarrow U_{\ell(l+k)}$ and giving the response $s = \mathsf{DEncode}(c, h, k, r)$. If $r \leftarrow U_{k \times lk}$, then $s \leftarrow D_1$, and if $r \leftarrow G(U_k)$, then $s \leftarrow D_2$. Thus A's advantage in distinguishing $G(U_k)$ and $U_{k \times lk}$ is exactly:

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prg}}_{A,G}(k) &= |\Pr[A(G(U_k)) = 1] - \Pr[A(U_{k \times lk}) = 1]| \\ &= \left|\Pr[W^{SD}(D_2) = 1] - \Pr[W^{SD}(D_1) = 1]\right| \\ &= \mathbf{Adv}^2_W(k)\end{aligned}$$

Lemma 3. $\mathbf{Adv}^3_W(k) \leq \mathbf{InSec}^{\mathsf{cca}}_{\mathcal{E}}(t', \boldsymbol{q}, \boldsymbol{\mu}, k) + \nu(k)$

Proof. We will construct an adversary A that plays the chosen-ciphertext attack game against $\mathcal{E}$ with advantage $\mathbf{Adv}^{\mathsf{cca}}_{A,\mathcal{E}}(k) \geq \mathbf{Adv}^3_W(k)$.

A starts by emulating W to get a challenge hiddentext, responding to decoding queries as follows: on query $(s_1, \ldots, s_l, h)$, A computes $c = f(s_1)\| \cdots \|f(s_l)$; A then uses its decryption oracle to compute $r\|_k m = D_{SK}(c)$. If $c \neq \perp$ and $s = \mathsf{DEncode}(c, h, k, G(r))$, A returns m, otherwise A returns $\perp$.

When W generates challenge (m^*, h^*), A chooses $r^* \leftarrow U_k$ and outputs the challenge $r^*\|m^*$. A is given the challenge ciphertext c^* and returns $s^* = \mathsf{DEncode}(c^*, h^*, k, G(r^*))$ to W.

A continues to emulate W, responding to queries as before, except that on decoding query $(s_1, \ldots, s_l, h)$, A first checks whether $f(s_1)\| \cdots \|f(s_l) = c^*$; if so, A returns $\perp$ rather than querying $D_{SK}(c^*)$.

In other words, A simulates running $\mathsf{SCCA.Decode}$ with its D_{SK} oracle, except that because A is playing the $\mathsf{IND\$\text{-}CCA}$ game, he is not allowed to query D_{SK} on the challenge value c^*: thus a decoding query that has the same underlying ciphertext c^* must be dealt with specially.

Notice that when A is given an encryption of $r^*\|m^*$, he perfectly simulates D_4 to W, so that $\Pr[A^{D_{SK}}(PK, E_{PK}(r^*\|m^*) = 1] = \Pr[W^{SD}(PK, D_4) = 1]$. This is because when $c^* = E_K(r^*\|m^*)$ then the test $s = \mathsf{DEncode}(c, h, k, G(r))$ would fail anyways. Likewise, when A is given a random string, he perfectly simulates D_3 to W, *given that c^* is not a valid ciphertext.* Let us denote the event that c^* is a valid ciphertext by V, and the event that a sample from D_3 encodes a valid ciphertext by U; notice that by construction $\Pr[\mathsf{U}] = \Pr[\mathsf{V}]$. We then have that

$$\begin{aligned}
\Pr[A^D(PK, U_\ell) = 1] &= \Pr[A^D(PK, U_\ell) = 1 | \overline{\mathsf{V}}] \Pr[\overline{\mathsf{V}}] + \Pr[A^D(PK, U_\ell) = 1 | \mathsf{V}] \Pr[\mathsf{V}] \\
&\leq \Pr[W^{SD}(PK, D_3) = 1 | \overline{\mathsf{U}}] \Pr[\overline{\mathsf{U}}] + \Pr[\mathsf{V}] \\
&\leq \Pr[W^{SD}(PK, D_3) = 1] + \Pr[\mathsf{V}] \\
&\leq \Pr[W^{SD}(PK, D_3) = 1] + \nu(k) \ ,
\end{aligned}$$

since $\Pr[\mathsf{V}] \leq \nu(k)$ by assumption on $\mathcal{E}$. Combining the cases, we find that

$$\begin{aligned}
\mathbf{Adv}^{\mathsf{cca}}_{A,\mathcal{E}}(k) &= \Pr[A^{D_{SK}}(PK, E_{PK}(r^*\|m^*) = 1] - \Pr[A^{D_{SK}}(PK, U_\ell) = 1] \\
&= \Pr[W^{SD}(PK, D_4) = 1] - \Pr[A^{D_{SK}}(PK, U_\ell) = 1] \\
&\geq \Pr[W^{SD}(PK, D_4) = 1] - \Pr[W^{SD}(PK, D_3) = 1] - \nu(k) \\
&= \mathbf{Adv}^3_W(k) - \nu(k)
\end{aligned}$$

Remark. As described, the stegosystem SCCA requires the decoder to know the algorithm $\mathsf{channel}$ used by the encoder to sample from $\mathcal{C}$. This can be avoided by changing the encoder to append a canonical encoding of this algorithm to the hiddentext before encrypting; the decoder then recovers this algorithm before running the final $\mathsf{DEncode}$ check. Since the length of the algorithm is constant, the security bounds for the resulting scheme are essentially unchanged.

6 Conclusion and open problems

We have argued for the importance of a $\mathsf{SS\text{-}CCA}$-secure stegosystem in the bare public key model, and given the first construction which meets this criterion. This resolves an open question posed by Backes and Cachin [3]. Furthermore, our construction relies on a public-key cryptosystem which is pseudorandom against chosen-ciphertext attack in the standard model. The existence of a cryptosystem satisfying this notion was an open problem posed by von Ahn and Hopper [1]. Because replay attacks are a realistic possibility, this represents an important advance over previous work.

One interesting direction for future work is to investigate the relationship between efficiently sampleable channels and the probabilistic channel oracle notion of earlier work. Designing a $\mathsf{SS\text{-}CCA}$ stegosystem in this setting seems to be a challenging problem. Another important notion of security against active attacks is *robustness* — the property that an attacker is unable to "remove" the hiddentext from a message. Hopper *et al* [13] define a weak notion of robustness and give a robust construction in the private key case. To our knowledge, there is

no provably secure construction satisfying this definition in the public-key case. It is interesting to note that SS-CCA and robustness are inherently contradictory, since robustness *requires* that a replay attack is possible. Thus it is also an interesting question whether some notion of robustness with decoding oracles can be achieved, even in the private key case.

References

1. L. von Ahn and N. Hopper. Public-Key Steganography. In: *Advances in Cryptology – Proceedings of Eurocrypt '04*, 2004.
2. R. J. Anderson and F. A. P. Petitcolas. On The Limits of Steganography. *IEEE Journal of Selected Areas in Communications*, 16(4), pages 474-481, 1998.
3. M. Backes and C. Cachin. Public-Key Steganography with Active Attacks. In: *Proc. Second Theory of Cryptography Conference (TCC)*, 2005.
4. M. Bellare and P. Rogaway. Random Oracles are Practical. In: *Proc. First ACM Conference on Computer and Communications Security (CCS 1993)*, 1993.
5. C. Cachin. An Information-theoretic model of steganography. In: *Information Hiding, 2nd International Workshop*, pages 306-318, 1998.
6. R. Canetti, H. Krawczyk, and J. Nielsen. Relaxing chosen-ciphertext security. In: *Advances in Cryptology – CRYPTO 2003*, 2003.
7. R. Cramer and V. Shoup. A practical public-key cryptosystem provably secure against adaptive chosen ciphertext attack. *Advances in Cryptology: CRYPTO 98*, Springer LNCS 1462, pages 13-27, 1998.
8. S. Craver. On Public-key Steganography in the Presence of an Active Warden. *Proceedings of Second International Information Hiding Workshop*, Springer LNCS 1525, pages 355-368, 1998.
9. N. Dedić, G. Itkis, L. Reyzin and S. Russell. Upper and lower bounds on black-box steganography. In: *Proc. Second Theory of Cryptography Conference (TCC)*, 2005.
10. D. Dolev and C. Dwork and M. Naor. Nonmalleable Cryptography. *SIAM J. Computing*, 30(2), pages 391–437, 2000.
11. R. Gennarro and V. Shoup. A Note on an Encryption Scheme of Kurosawa and Desmedt. *IACR e-print archive report 2004/194*, 2004.
12. O. Goldreich. *Foundations of Cryptography: volume 1 – Basic Tools.* Cambridge University Press, 2001.
13. N. J. Hopper, J. Langford, and L. Von Ahn. Provably Secure Steganography. In: *Advances in Cryptology – CRYPTO 2002*, Springer LNCS 2442, pages 77-92, 2002.
14. N.J. Hopper. Toward a theory of steganography. Ph.D. Thesis, Carnegie Mellon University, July 2004. Available online: http://reports-archive.adm.cs.cmu.edu/anon/2004/abstracts/04-157.html
15. S. Katzenbeisser and F. A. P. Petitcolas. Defining Security in Steganographic Systems. In: *Proceedings of the SPIE vol. 4675, Security and Watermarking of Multimedia Contents IV*, pp. 50-56, 2002.
16. K. Kurosawa and Y. Desmedt. A New Paradigm of Hybrid Encryption Scheme. In: *Advances in Cryptology – Proceedings of CRYPTO '04*, 2004.
17. T. V. Le and K. Kurosawa. Efficient public key steganography secure against adaptive chosen stegotext attacks. *IACR e-print archive report 2003/244*, 2003.
18. A. Lysyanskaya and M. Meyerovich. Steganography with imperfect sampling. At: *CRYPTO 2004 Rump Session*, August 2004.
19. B. Möller. A Public-Key Encryption Scheme with Pseudorandom Ciphertexts. In: *Computer Security – ESORICS 2004*, 2004.