# Efficient Constructions of Deterministic Encryption from Hybrid Encryption and Code-based PKE[*]

Yang Cui[1], Kirill Morozov[2], Kazukuni Kobara[3], and Hideki Imai[4]
*(Corresponding author: Kirill Morozov)*

Wireless Network Research Department, Huawei Technologies Co. Ltd.[1]
No.156, BeiQing Rd., Haidian-District, Beijing 100095 China
Institute of Mathematics for Industry, Kyushu University[2]
744 Motooka, Nishi-ku, Fukuoka 819-0395 Japan
Research Institute for Secure Systems, National Institute of Advanced Industrial Science and Technology[3]
1-1-1 Umezono, Tsukuba-shi, Ibaraki-ken 305-0046 Japan
Department of Electrical, Electronic and Communication Engineering, Chuo University[4]
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551 Japan
(Email: cuilang@gmail.com, morozov@imi.kyushu-u.ac.jp, k-kobara@aist.go.jp, h-imai@imailab.jp)

*(Received Feb. 28, 2012; revised and accepted Aug. 6, 2012)*

## Abstract

We present efficient constructions of deterministic encryption (DE) satisfying the new notion – security against privacy adversary (PRIV security), in the random oracle model. Our work includes: 1) A generic construction of deterministic length-preserving hybrid encryption, which is an improvement over the paper by Bellare et al. in Crypto'07; to our best knowledge, this is the first example of length-preserving deterministic hybrid encryption (DHE); 2) post-quantum deterministic encryption, using the code-based encryption, which enjoys a simplified construction since its public key is re-used as a hash function; 3) deterministic encryption with high message rate from witness-recovering encryption.

*Keywords: code-based encryption, database security, deterministic encryption, hybrid encryption, searchable encryption*

## 1 Introduction

### 1.1 Background

The notion of security against *privacy adversary* (denoted as PRIV) for deterministic encryption (DE) was pioneered by Bellare et al. [2] featuring an upgrade from the standard one-wayness property. Instead of not leaking the whole plaintext, the ciphertext was demanded to leak, roughly speaking, no more than the plaintext statistics does. In other words, the PRIV-security definition (formulated in a manner similar to the semantic security definition of [9]) requires that a ciphertext must be essentially useless for adversary who is to compute some predicate on the corresponding plaintext. Achieving PRIV-security demands two important assumptions: 1) the plaintext space must be large enough and must have a smooth (i.e. high min-entropy) distribution; 2) the plaintext and the predicate are independent of the public key.

Constructions satisfying two flavors of PRIV-security are presented in [2]: against chosen-plaintext (CPA) and chosen-ciphertext (CCA) attacks. The following three PRIV-CPA constructions are introduced in the random oracle (RO) model. The generic Encrypt-with-Hash (EwH) primitive features replacing the coins used by the randomized encryption scheme with a hash of the public key concatenated with the message. The RSA deterministic OAEP (RSA-DOAEP) scheme provides us with length-preserving DE. In the generic Encrypt-and-Hash (EaH) primitive, a "tag" in the form of the plaintext's hash is attached to the ciphertext of a randomized encryption scheme.

These results were extended by Boldyreva et al. [5] and Bellare et al. [3] presenting new extended definitions, proving relations between them, and introducing, among

[*] An extended abstract of this paper was presented at the Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, 18th International Symposium, AAECC-18 2009, Tarragona, Catalonia, Spain, June 8-12, 2009.

others, new constructions without random oracles.

### 1.2 Applications

The original motivation for this research comes from the demand on efficiently searchable encryption (ESE) in the database applications. Length-preserving schemes can also be used for encryption of legacy code and in the bandwidth-limited systems. Some more applications (although irrelevant to our work) to improving randomized encryption schemes were studied in [5].

### 1.3 Motivation

The work [2] sketches a method for encrypting long messages, but it is less efficient compared to the standard hybrid encryption, besides it is conjectured not to be length-preserving. Also, possible emerging of quantum computers raises demands for post-quantum DE schemes.

### 1.4 Our Contribution

In this work, we assume existence of idealized hash functions which behave like random oracles, i.e. our results are in the random oracle model [4]. We present a generic and efficient construction of length-preserving deterministic hybrid encryption (DHE). In a nutshell, we prove that the session key can be computed by concatenating the public key with the first message block and inputting the result into key derivation function. This is a kind of re-using the (sufficient) entropy of message, and it is secure due to our assumption that the first block of the message is of high min-entropy and independent of the key. In a sense, we buy the length preserving property for the price of restricting the plaintext distribution. This assumption is meaningful in some practical contexts: for instance, in a telephone database, the area code may be fixed, while the individual number is highly unpredictable.

Compared to our case, Bellare et al. employ the hybrid encryption in a conventional way, which first encrypts a random session key to further encrypt the data, obviously losing the length-preserving property. Hence, we show that the claim of Bellare et al. [2]: “However, if using hybrid encryption, RSA-DOAEP would no longer be length-preserving (since an encrypted symmetric key would need to be included with the ciphertext)” is overly pessimistic. To our best knowledge, this is the first example of length-preserving hybrid encryption.

For achieving post-quantum DE, we propose to plug in an IND-CPA secure variant [11] of the coding theory based (or code-based) McEliece public key encryption (PKE) [10] into the generic constructions EaH and EwH, presented in [2]. The McEliece PKE is believed to be resistant to quantum attacks, besides it has very fast encryption algorithm. Moreover, we point out a significant simplification: the public key (which is a generating matrix of some linear code) can be re-used as hash function.

In witness-recovering encryption, one decodes from the ciphertext not only the plaintext, but also the random coin (witness) which is used to generate the ciphertext. We show that such schemes can be used to construct DE with longer plaintext (as compared to the original schemes). The idea is to have the witness carry additional information, while preserving security of the scheme. For the same reason as in the DHE construction, we require that the first block of the message is of high min-entropy and independent of the key.

### 1.5 Related Work

A deterministic hybrid encryption scheme was proposed in the RSA-DOAEP scheme of [2]. Our proposal uses the same principle, but we provide a generic construction, which works for particular message distributions. There are several recent work on DE, such as [3, 5], which prove security in the standard model (without the help of random oracles). However, their constructions are somewhat inefficient with the sole exception of the scheme [3] based on the Decisional Composite Residuosity assumption.

### 1.6 Organization

The paper will be organized in the following way: Section 2 provides the security definitions for DE. Section 3 gives the proposed generic and efficient construction of DHE, which immediately leads to the first length-preserving construction. In Section 4, we will provide DE from the code-based PKE, which is post-quantum secure and efficient due to the good property of the underlying PKE scheme. Next, in Section 5, on observing that many code-based PKE are also witness-recovering encryption at the same time, we propose a high message rate DE tailored to it. In Section 6, we briefly discuss how to extend security of our schemes to the chosen-ciphertext attack (CCA) scenario. Finally, we provide concluding remarks in Section 7.

## 2 Preliminaries

Denote by “$|x|$” the cardinality of $x$. Denote by $\vec{x}$ the vector and by $\vec{x}[i]$ the $i$-th component of $\vec{x}$ ($1 \leq i \leq |\vec{x}|$). Write $\vec{x}||\vec{y}$ for concatenation of vectors $\vec{x}$ and $\vec{y}$. Let $x \leftarrow_R X$ denote the operation of picking $x$ from the set $X$ uniformly at random. Denote by $z \leftarrow A(x, y, ...)$ the operation of running algorithm $A$ with input $(x, y, ...)$, to output $z$. Write $\log x$ as the logarithm with base 2. We also write $\Pr[A(x) = y : x \leftarrow_R X]$ the probability that $A$ outputs $y$ corresponding to input $x$, which is sampled from $X$. We say a function $\epsilon(k)$ is negligible, if for any constant $c$, there exists $k_0 \in \mathbb{N}$, such that $\epsilon < (1/k)^c$ for any $k > k_0$.

A public key encryption (PKE) scheme $\Pi$ consists of a triple of algorithms $(\mathcal{K}, \mathcal{E}, \mathcal{D})$. The key generation algorithm $\mathcal{K}$ outputs a pair of public and secret keys $(\mathsf{pk}, \mathsf{sk})$

taking on input $1^k$, a security parameter $k$ in unitary notation. The encryption algorithm $\mathcal{E}$ on input $\mathsf{pk}$ and a plaintext $\vec{x}$ outputs a ciphertext $c$. The decryption algorithm $\mathcal{D}$ takes $\mathsf{sk}$ and $c$ as input and outputs the plaintext message $\vec{x}$. We require that for any key pair $(\mathsf{pk}, \mathsf{sk})$ obtained from $\mathcal{K}$, and any plaintext $\vec{x}$ from the plaintext space of $\Pi$, $\vec{x} \leftarrow \mathcal{D}(\mathsf{sk}, \mathcal{E}(\mathsf{pk}, \vec{x}))$.

**Definition 1 (PRIV [2]).** *Let a probabilistic polynomial-time (PPT) adversary $\mathcal{A}_{DE}$ against the privacy of the DE scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, be a pair of algorithms $\mathcal{A}_{DE} = (\mathcal{A}_f, \mathcal{A}_g)$, where $\mathcal{A}_f, \mathcal{A}_g$ do not share any random coins or state. The advantage of adversary is defined as follows,*

$$\begin{aligned}\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}_{DE}}(k) = \Pr[\mathbf{Exp}^{priv-1}_{\Pi,\mathcal{A}_{DE}}(k) = 1] \\ - \Pr[\mathbf{Exp}^{priv-0}_{\Pi,\mathcal{A}_{DE}}(k) = 1],\end{aligned}$$

*where experiments are described as:*

$$\begin{array}{l|l}
\mathbf{Exp}^{priv-1}_{\Pi,\mathcal{A}_{DE}}(k): & \mathbf{Exp}^{priv-0}_{\Pi,\mathcal{A}_{DE}}(k): \\
 & \\
(\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}(1^k) & (\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}(1^k) \\
(\vec{x}_1, t_1) \leftarrow_R \mathcal{A}_f(1^k) & (\vec{x}_0, t_0) \leftarrow_R \mathcal{A}_f(1^k) \\
 & (\vec{x}_1, t_1) \leftarrow_R \mathcal{A}_f(1^k) \\
c \leftarrow_R \mathcal{E}(1^k, \mathsf{pk}, \vec{x}_1) & c \leftarrow_R \mathcal{E}(1^k, \mathsf{pk}, \vec{x}_0) \\
g \leftarrow_R \mathcal{A}_g(1^k, \mathsf{pk}, c) & g \leftarrow_R \mathcal{A}_g(1^k, \mathsf{pk}, c) \\
\textit{Return 1 if } g = t_1 & \textit{Return 1 if } g = t_1 \\
\textit{Else return 0} & \textit{Else return 0}
\end{array}$$

*We say that* $\Pi$ *is* PRIV secure, *if* $\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}_{DE}}(k)$ *is negligible, for any PPT $\mathcal{A}_{DE}$ with high min-entropy, where $\mathcal{A}_{DE}$ has a high min-entropy $\mu(k)$ means that $\mu(k) \in \omega(log(k))$, and $\Pr[\vec{x}[i] = x : (\vec{x}, t) \leftarrow_R \mathcal{A}_m(1^k)] \leq 2^{-\mu(k)}$ for all $k$, all $1 \leq i \leq |\vec{x}|$, and any $x \in \{0,1\}^*$.*

In the underlying definition, the advantage of privacy adversary could be also written as

$$\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}_{DE}}(k) = 2\Pr[\mathbf{Exp}^{priv-b}_{\Pi,\mathcal{A}_{DE}}(k) = b] - 1$$

where $b \in \{0,1\}$ and probability is taken over the choice of all of the random coins in the experiments.
*Remarks.*

1) The encryption algorithm $\Pi$ need not be deterministic per se. For example, in a randomized encryption scheme, the random coins can be fixed in an appropriate way to yield a deterministic scheme (as explained in Section 4);

2) As argued in [2], $\mathcal{A}_f$ has no access to $\mathsf{pk}$ and $\mathcal{A}_g$ does not know the chosen plaintext input to encryption oracle by $\mathcal{A}_f$. This is required because the public key itself carries some non-trivial information about the plaintext if the encryption is deterministic.[1] Thus, equipping either $\mathcal{A}_f$ or $\mathcal{A}_g$ with both the public key and free choice of an input plaintext in the way of conventional indistinguishability notion [9] of PKE, the PRIV security cannot be achieved.

It is possible to build PRIV security from indistinguishability (IND) security, as observed in [2]. In the following, we recall the notion of IND security.

**Definition 2 (IND-CPA).** *We say a scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is IND-CPA secure, if the advantage $\mathbf{Adv}^{ind}_{\Pi,\mathcal{A}}$ of any PPT adversary $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ is negligible, (let $s$ be the state information of $\mathcal{A}_1$, and $\hat{b} \in \{0,1\}$):*

$$\mathbf{Adv}^{ind}_{\Pi,\mathcal{A}}(k) = 2 \cdot \Pr\left[\begin{array}{l} \hat{b} = b : (\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}(1^k), \\ (x_0, x_1, s) \leftarrow_R \mathcal{A}_1(1^k, \mathsf{pk}), \\ b \leftarrow_R \{0,1\}, c \leftarrow_R \mathcal{E}(1^k, \mathsf{pk}, x_b), \\ \hat{b} \leftarrow_R \mathcal{A}_2(1^k, c, s) \end{array}\right] - 1$$

*Remark.* IND security is required by a variety of cryptographic primitives. However, for an efficiently searchable encryption used in database applications, IND secure encryption may be considered as overkill. For such a strong encryption, it is not known how to arrange fast (i.e. logarithmic in the database size) search.

IND secure symmetric key encryption (SKE) has been carefully discussed in the literature, such as [7]. Given a key $K \in \{0,1\}^k$ and message $m$, an encryption algorithm outputs a ciphertext $\chi$. Provided $\chi$ and $K$, a decryption algorithm outputs the message $m$ uniquely. Note that for a secure SKE, outputs of the encryption algorithm could be considered uniformly distributed in the range, when encrypted under independent session keys. Besides, IND secure SKE is easy to construct.

**Definition 3 (IND-CPA SKE).** *A symmetric key encryption scheme denoted as $\Lambda = (\mathcal{K}_{SK}, \mathcal{E}_{SK}, \mathcal{D}_{SK})$ with key space $\{0,1\}^k$, is indistinguishable against chosen plaintext attack (IND-CPA) if the advantage of any PPT adversary $\mathcal{B}$, $\mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}$ is negligible, where*

$$\mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}(k) = 2 \cdot \Pr\left[\begin{array}{l} \hat{b} = b : K \leftarrow_R \{0,1\}^k, \\ b \leftarrow_R \{0,1\}, \\ \hat{b} \leftarrow_R \mathcal{B}^{\mathsf{LOR}(K,\cdot,\cdot,b)}(1^k) \end{array}\right] - 1,$$

*where a left-or-right oracle $\mathsf{LOR}(K, M_0, M_1, b)$ returns $\chi \leftarrow_R \mathcal{E}_{SK}(K, M_b)$. Adversary $\mathcal{B}$ is allowed to ask $\mathsf{LOR}$ oracle, with two chosen message $M_0$, $M_1$ ($M_0 \neq M_1$, $|M_0| = |M_1|$).*

**Hybrid Encryption.** In the seminal paper by Cramer and Shoup [7], the idea of hybrid encryption is rigorously studied. Note that typically, PKE is applied in key distribution process due to its high computational cost, while SKE is typically used for encrypting massive data flow using a freshly generated key for each new session. In hybrid encryption, PKE and SKE work in tandem: a randomly generated session key is first encrypted by PKE, then the plaintext is further encrypted on the session key by SKE. Hybrid encryption is more

[1] In other words, suppose that in Def. 1, $\mathcal{A}_f$ knows $\mathsf{pk}$. Then, $\mathcal{A}_f$ can assign $t_1$ to be the ciphertext $c$, and hence $\mathcal{A}_g$ always wins the game (returns 1). Put it differently, although $\mathcal{A}_f$ and $\mathcal{A}_g$ are not allowed to share a state, the knowledge of $\mathsf{pk}$ can help them to share it anyway.

Table 1: Generic construction of deterministic hybrid encryption

| $\mathcal{K}_H(1^k)$: | $\mathcal{E}_H(\mathsf{pk}, x)$: | $\mathcal{D}_H(\mathsf{sk}, c)$: |
|---|---|---|
| $(\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}(1^k)$ | Parses $x$ to $\bar{x}\Vert\underline{x}$ | Parse $c$ to $\psi\Vert\chi$ |
| Return $(\mathsf{pk}, \mathsf{sk})$ | $\psi \leftarrow_R \mathcal{E}(1^k, \mathsf{pk}, \bar{x})$ | $\bar{x} \leftarrow \mathcal{D}(\mathsf{sk}, \psi)$ |
| | $K \leftarrow H(\mathsf{pk}\Vert\bar{x})$ | $K \leftarrow H(\mathsf{pk}\Vert\bar{x})$ |
| | $\chi \leftarrow_R \mathcal{E}_{SK}(K, \underline{x})$ | $\underline{x} \leftarrow \mathcal{D}_{SK}(K, \chi)$ |
| | Return $c = \psi\Vert\chi$ | Return $x = \bar{x}\Vert\underline{x}$ |

commonly used in practice than a sole PKE, since encryption/decryption of the former is substantially faster for long messages.

**Deterministic Hybrid Encryption.** A deterministic public-key encryption could be easily extended to the hybrid scenario, in addition to a SKE. Actually, as [2] argued, a deterministic SKE is easier to define and achieve, in the left-or-right oracle model, where the challenge messages are distinct. Hence, for obtaining a secure DHE, we simply require both PKE and SKE to be PRIV secure.

# 3 Secure Deterministic Hybrid Encryption

In this section, we will present a generic composition of PKE and SKE to obtain DHE. Interestingly, the our result is quite different from conventional hybrid encryption. In that case, the overhead of communication cost includes at least the size of the session key, even if we pick the PKE scheme being a (length-preserving) one-way trapdoor permutation, e.g. RSA.

However, we notice that in the PRIV security definition, both the public key and the plaintext are not simultaneously known by either $\mathcal{A}_f$ or $\mathcal{A}_g$. Hence, one can save on generating and encrypting a random session key. Instead, the secret session key could be extracted from the combination of public key and plaintext which are available to a legal user contrary to the adversary.

## 3.1 Generic Composition of PRIV-secure PKE and IND-CPA Symmetric Key Encryption

Given a PRIV secure PKE scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, and an IND-CPA secure SKE scheme $\Lambda = (\mathcal{K}_{SK}, \mathcal{E}_{SK}, \mathcal{D}_{SK})$, we can achieve a deterministic hybrid encryption scheme $DHE = (\mathcal{K}_H, \mathcal{E}_H, \mathcal{D}_H)$. In the following, $H : \{0,1\}^* \mapsto \{0,1\}^k$ is a key derivation function (KDF), modeled as a random oracle. In the following section, we simply write input vector $\vec{x}$ as $x$ with length of $|\vec{x}| = v$. Wlog, parse $x = \bar{x}||\underline{x}$, where the $|\bar{x}|$ and $|\underline{x}|$ is the size (in bits) of the input domain of $\Pi$ and $\Lambda$, respectively.

Our proposed construction is presented in Table 1.

It is simple, efficient, and can be generically built from any PRIV $\Pi$ and IND-CPA $\Lambda$. Note that the secret session key must have high min-entropy in order to deny a brute-force attack against $\Lambda$. The high min-entropy requirement should be fulfilled for any PPT privacy adversary to $\Pi$ since otherwise, PRIV security is not available, as pointed out in [2]. Thus, we can build a reduction of security of DHE to that of deterministic PKE.

Requiring $\bar{x}$ to be of high min-entropy, rules out a trivial attack, which can be described by the following example. Suppose that a DHE's input $x = \bar{x}||\underline{x}$, where $\bar{x}$ is fixed to a certain number, say all zero. $\mathcal{A}_f$ outputs $0\ldots0||\underline{x}$ and sets $t = \underline{x}$. Even though $\underline{x}$ may have high min-entropy $\mu(k)$, adversary $\mathcal{A}_g$ can compute $K = H(pk||0\ldots0)$, and thus decrypt $\underline{x}$ from $\chi$ with $K$. $\mathcal{A}_g$ can always successfully output $g = \underline{x}$, which is equivalent to $t$. This attack works since the input $\bar{x}$ to $\Pi$ has a very low min-entropy, that, in particular, does not satisfy the conditions of PRIV security of $\Pi$.

As we have explained, for preventing such a trivial attack, we set a high min-entropy requirement of adversary to PRIV $\Pi$. Note, however, that we did not set any restrictions on the $\underline{x}$ – even a fixed one will yield a secure scheme.

Next, we will provide our security proof of proposed DHE.

## 3.2 Security Proof

**Theorem 1.** *In the random oracle model, given a PRIV PKE scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, and an IND-CPA SKE scheme $\Lambda = (\mathcal{K}_{SK}, \mathcal{E}_{SK}, \mathcal{D}_{SK})$, if there is a PRIV adversary $\mathcal{A}_H$ against the hybrid encryption $DHE = (\mathcal{K}_H, \mathcal{E}_H, \mathcal{D}_H)$, then there exists a PRIV adversary $\mathcal{A}$ and an IND-CPA adversary $\mathcal{B}$, such that*

$$\mathbf{Adv}^{priv}_{DHE,\mathcal{A}_H}(k) \leq \mathbf{Adv}^{priv}_{\Pi,\mathcal{A}}(k) + \mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}(k) + q_h v/2^{\mu}$$

*where $q_h$ is an upper bound on the number of queries to the random oracle $H$, $v$ is the plaintext size of $\Pi$, $\mu$ is defined by high min-entropy of PRIV security of $\Pi$.*

*Proof.* We will provide the security proof in the game-hopping way, namely start from a PRIV adversary $\mathcal{A}_H = (\mathcal{A}_f, \mathcal{A}_g)$ to DHE scheme in experiment $\mathbf{Exp}^{priv-1}_{DHE,\mathcal{A}_H}(k)$, and gradually modify the game so that we can obtain similar result in experiment $\mathbf{Exp}^{priv-0}_{DHE,\mathcal{A}_H}(k)$, otherwise we can build PPT adversary $\mathcal{A}$ to break PRIV security of $\Pi$ and $\mathcal{B}$ to break IND-CPA security of $\Lambda$.

The original game for PRIV security of DHE is shown in Figure 1. □

More precisely, if a successful adversary for this game exists, then

$$\begin{aligned}\mathbf{Adv}^{priv}_{DHE,\mathcal{A}_H}(k) = \Pr[\mathbf{Exp}^{priv-1}_{DHE,\mathcal{A}_H}(k) = 1] \\ - \Pr[\mathbf{Exp}^{priv-0}_{DHE,\mathcal{A}_H}(k) = 1]\end{aligned}$$

is non-negligible for some $\mathcal{A}_H$. Next we present a simulator which gradually modifies the above experiments such that the adversary does not notice it. Our goal is

```
$\mathbf{Exp}^{priv-1}_{DHE,\mathcal{A}_H}(k)$:

$(\mathsf{pk},\mathsf{sk}) \leftarrow_R \mathcal{K}(1^k)$
$(x_1,t_1) \leftarrow_R \mathcal{A}_f(1^k)$

Parse $x_1$ to $\bar{x}_1||\underline{x}_1$
$\psi \leftarrow_R \mathcal{E}(1^k,\mathsf{pk},\bar{x}_1)$
$K \leftarrow H(\mathsf{pk}||\bar{x}_1)$
$\chi \leftarrow_R \mathcal{E}_{SK}(K,\underline{x}_1)$
$c \leftarrow \psi||\chi$
$g \leftarrow_R \mathcal{A}_g(1^k,\mathsf{pk},c)$
Return 1 if $g = t_1$
Else return 0
```

```
$\mathbf{Exp}^{priv-0}_{DHE,\mathcal{A}_H}(k)$:

$(\mathsf{pk},\mathsf{sk}) \leftarrow_R \mathcal{K}(1^k)$
$(x_0,t_0) \leftarrow_R \mathcal{A}_f(1^k)$
$(x_1,t_1) \leftarrow_R \mathcal{A}_f(1^k)$
Parse $x_0$ to $\bar{x}_0||\underline{x}_0$
$\psi' \leftarrow_R \mathcal{E}(1^k,\mathsf{pk},\bar{x}_0)$
$K' \leftarrow H(\mathsf{pk}||\bar{x}_0)$
$\chi' \leftarrow_R \mathcal{E}_{SK}(K',\underline{x}_0)$
$c' \leftarrow \psi'||\chi'$
$g \leftarrow_R \mathcal{A}_g(1^k,\mathsf{pk},c')$
Return 1 if $g = t_1$
Else return 0
```

Figure 1: The original game for PRIV security of DHE

to show that $\mathbf{Adv}^{priv}_{DHE,\mathcal{A}_H}(k)$ is almost as big as the corresponding advantages defined for PRIV security of the PKE scheme and IND-CPA security of the SKE scheme, which are assumed negligible.

Because of the high min-entropy requirement of PRIV adversary, it is easy to see that $x_0 \neq x_1$, except with negligible probability. Thus, we disregard the above possibility and consider the following cases: $\bar{x}_0 \neq \bar{x}_1$, or $\underline{x}_0 \neq \underline{x}_1$, or both.

**Case** $[\bar{x}_0 \neq \bar{x}_1]$ Since $x_0 \neq x_1$ and $\bar{x}_0 \neq \bar{x}_1$, the right part of $x_b$ $(b \in \{0,1\})$, could be equal or not.

- When $\underline{x}_0 = \underline{x}_1$, the adversary has two targets, such as $\Pi$ and $\Lambda$ in two experiments. First look at the SKE scheme $\Lambda$. In this case, the inputs to $\Lambda$ in two experiments are the same, but still unknown to $\mathcal{A}_g$. The key derivation function $H$ outputs $K \leftarrow H(\mathsf{pk}||\bar{x}_1)$ and $K' \leftarrow H(\mathsf{pk}||\bar{x}_0)$. Since $\bar{x}_0 \neq \bar{x}_1$, we have $K \neq K'$. Note that $\mathcal{A}_g$ does not know $x_0$ nor $x_1$, thus does not know $K, K'$, either. Then, $\mathcal{A}_g$ must tell which of $\chi, \chi'$ is the corresponding encryption under the unknown keys without knowing $\underline{x}_0, \underline{x}_1 (\underline{x}_0 = \underline{x}_1)$, which is harder than breaking IND-CPA security and that could be bounded by $\mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}(k)$.

  On the other hand, the adversary can also challenge the PKE scheme $\Pi$ to distinguish two experiments, but it will break the PRIV security. More precisely, the advantage in distinguishing $\psi, \psi'$ with certain $K, K'$ is at most $\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}}(k)$, since $K, K'$ are not output explicitly and unavailable to adversary.

- When $\underline{x}_0 \neq \underline{x}_1$, this case is similar to the above, except that the inputs to $\Lambda$ are different. $\mathcal{A}_g$ can do nothing given $\chi, \chi'$ only, hence $\mathcal{A}_g$'s possible attack must be focused on $\Pi$, and its advantage can be bounded by $\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}}(k)$.

**Case** $[\underline{x}_0 \neq \underline{x}_1]$ Similarly, there must be either $\bar{x}_0 \neq \bar{x}_1$ or $\bar{x}_0 = \bar{x}_1$.

- When $\bar{x}_0 = \bar{x}_1$, the same session key $K \leftarrow H(\mathsf{pk}||\bar{x}_b)$ $(b \in \{0,1\})$ is used for $\Lambda$. In this case, the ciphertexts $\psi, \psi'$ are the same, adversary will focus on distinguishing $\chi, \chi'$. Note that $\mathcal{A}_f$ cannot compute $K$ even though he knows $\bar{x}_0$ (or equivalently $\bar{x}_1$), because $\mathsf{pk}$ is not known to him (otherwise, it will break the PRIV security of $\Pi$ immediately!). Thus, the successful distinguishing requires $\mathcal{A}_g$ to choose the same $\bar{x}_0 = \bar{x}_1$ when querying to the random oracle. Then, $\mathcal{A}_g$ has a harder game than IND-CPA (because it does not know $\underline{x}_0, \underline{x}_1$), whose advantage is bounded by $\mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}(k)$.

  In order to be sure that adversary $(\mathcal{A}_f, \mathcal{A}_g)$ mounting a brute-force attack to find out the session key of $\Lambda$ cannot succeed, the probability to find the key in searching all the random oracle queries should be taken into account as well. Suppose that adversary makes at most $q_h$ queries to its random oracle, and the $\Pi$'s plaintext size is $v$. Then, this probability could be upper bounded by $q_h v/2^{\mu}$ (Note that this bound is in nature similar to that in [2]).

- When $\bar{x}_0 \neq \bar{x}_1$, as we have discussed above, this will break the PRIV security of $\Pi$, and advantage of adversary could be bounded by $\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}}(k)$.

Summarizing, we conclude that in all cases when $(\mathcal{A}_f, \mathcal{A}_g)$ intends to break the PRIV security of our DHE scheme, its advantage of distinguishing two experiments is bounded by the sum of $\mathbf{Adv}^{priv}_{\Pi,\mathcal{A}}(k)$, $q_h v/2^{\mu}$ and $\mathbf{Adv}^{ind-cpa}_{\Lambda,\mathcal{B}}(k)$.

### 3.3 Length-preserving Deterministic Hybrid Encryption

The first length-preserving PRIV PKE scheme is RSA-DOAEP due to [2]. The length-preserving property is important in practice, for instance in bandwidth-restricted applications. RSA-DOAEP makes use of the RSA trapdoor permutation and with a modified 3-round Feistel network achieves the same sizes of input and output, i.e. $|m_{DE}| = |c_{DE}|$. As we have proved in Theorem 1, a construction proposed in Table 1 leads to a DHE.

Denote a length-preserving DHE that is composed of DE and SKE, s.t. $|m_{DE}| + |m_{SKE}| = |c_{DE}| + |c_{SKE}|$. In particular, RSA-DOAEP + IND-CPA SKE $\Rightarrow$ a length-preserving DHE, because both RSA-DOAEP and IND-CPA SKE are length-preserving. Note that in [2], it is argued that RSA-DOAEP based hybrid encryption scheme cannot be length-preserving any more, because a random session key has to be embedded in RSA-DOAEP. However, by re-using the knowledge of public key $\mathsf{pk}$ and a part of the message, we can indeed build the first length-preserving DHE, which is not only convenient in practice, but also meaningful in theory.

### 3.3.1 Security Proof

According to Theorem 1, a PRIV PKE scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ and an IND-CPA SKE scheme $\Lambda = (\mathcal{K}_{SK}, \mathcal{E}_{SK}, \mathcal{D}_{SK})$ suffice to construct a PRIV hybrid encryption $DHE = (\mathcal{K}_H, \mathcal{E}_H, \mathcal{D}_H)$. Besides, RSA-DOAEP is length-preserving and PRIV secure according to [2].

**Corollary 1.** *Denote by $\Lambda$ any IND-CPA SKE scheme with its input length equal to output length. $DHE = (\mathcal{K}_H,\ \mathcal{E}_H,\ \mathcal{D}_H)$ composed of $\Pi = (\mathcal{K}_{DOAEP},\ \mathcal{E}_{DOAEP},\ \mathcal{D}_{DOAEP})$ and $\Lambda = (\mathcal{K}_{SK},\ \mathcal{E}_{SK},\ \mathcal{D}_{SK})$ is PRIV secure and length-preserving.*

*Proof.* It is concluded directly from Theorem 1 and [2], since both RSA-DOAEP and $\Lambda$ are length-preserving. □

## 4 Deterministic Encryption from Code-based PKE

From a post-quantum point of view, it is desirable to obtain DE based on assumptions other than RSA or discrete log. Code-based PKE, such as McEliece PKE [10] is considered a promising candidate after being carefully studied for over thirty years.

McEliece PKE. Denoted $\Pi_M = (\mathcal{K}_M, \mathcal{E}_M, \mathcal{D}_M)$, i.e. it consists of the following triple of algorithms [10].

1) Key generation $\mathcal{K}_M$: On input a security parameter $\lambda$, output $(\mathsf{pk}, \mathsf{sk})$ as follows, such that $n, t \in \mathbb{N}$, $t \ll n$.
   - $\mathsf{sk}$ (Private Key): $(S, \varphi, P)$
     $G'$: $l \times n$ generating matrix of a binary irreducible $[n, l]$ Goppa code which can correct a maximum of $t$ errors, $\varphi$: an efficient bounded distance decoding algorithm of the underlying code, S: $l \times l$ non-singular matrix, P: $n \times n$ permutation matrix, chosen at random.
   - $\mathsf{pk}$ (Public Key): $(G, t)$
     G: $l \times n$ matrix given by a product of three matrices $SG'P$.

2) Encryption $\mathcal{E}_M$: Given $\mathsf{pk}$ and an $l$-bit plaintext $m$, randomly generate $n$-bit $e$ with Hamming weight $t$, output ciphertext $c = mG \oplus e$.

3) Decryption $\mathcal{D}_M$: On input $c$, output $m$ given $\mathsf{sk}$.
   - Compute $cP^{-1} = (mS)G' \oplus eP^{-1}$, where $P^{-1}$ is an inverse matrix of $P$.
   - Error correcting algorithm $\varphi$ corresponding to $G'$ applies to compute $mS = \varphi(cP^{-1})$.
   - Compute the plaintext $m = (mS)S^{-1}$.

IND-CPA security of the McEliece PKE can be achieved by padding the plaintext with a random bitstring $r$, $|r| = \lceil a \cdot l \rceil$ for some $0 < a < 1$. We refer to [11] for details.

Postquantum security is not the only motivation to achieve DE from code-based PKE. Another good property of the McEliece PKE and its variants is that its public key (being a generating matrix of an error-correcting code) could be used as a hash function to digest the message. The fact that a hash function can be based on hardness of decoding was originally noted by Stern [13]. Recently, such the function was designed and studied in [1, 8]. The advantage that public key itself is able to work as a hash function, can do us a favor to build efficient post-quantum DE. We call this Hidden Hash (HH) property of McEliece PKE.[2] Henceforth, we assume that this function behaves as a random oracle.

In [2], two constructions satisfying PRIV security have been proposed: Encrypt-with-Hash (EwH) and Encrypt-and-Hash (EaH). Adapting the HH property of the McEliece PKE to the both constructions, we can achieve PRIV secure DE. For proving PRIV security, we require the McEliece PKE to be IND-CPA secure, which has been achieved in [11].

**Construction of EwH.** Let $\Pi_M = (\mathcal{K}_M, \mathcal{E}_M, \mathcal{D}_M)$ be the IND-CPA McEliece PKE as described in Section 2, based on $[n, l, 2t+1]$ Goppa code family, with $l_p$-bit padding where $l_p = \lceil a \cdot l \rceil$ for some $0 < a < 1$, and plaintext length $l_m = l - l_p$. Let $\mathcal{H}$ be a hash family defined over a set of public keys of the McEliece PKE. $H_M : \{0,1\}^{l_m} \mapsto \{0,1\}^{l_p + \lceil \log \sum_{i=1}^{t} \binom{n}{t} \rceil}$ and $H_N : \{0,1\}^{l_m} \mapsto \{0,1\}^{2k}$ are uniquely defined by $1^k$ and $\mathsf{pk}$. Without knowledge of $\mathsf{pk}$, there is no way to compute $H_M$ or $H_N$ (refer to [1, 8] for details). Let $e$ be an error vector, s.t. $|e| = n$ with Hamming weight $Hw(e) = t$. According to Cover's paper [6], it is quite efficient to find an injective mapping to encode the bit string $r_e$ of length $\lceil \log \sum_{i=1}^{t} \binom{n}{t} \rceil$ into $e$, and vice versa.

Our EwH scheme is presented in Table 2. Note that compared with the EwH scheme proposed by Bellare et al. [2], our scheme does not need to include $\mathsf{pk}$ into the hash, because hash function $H_M$ itself is made of $\mathsf{pk}$. Public key $\mathsf{pk}$ could be considered as a part of the algorithm of the hash function, as well. When we model $H_M$ as a random oracle, we can easily prove the PRIV security in a similar way as Bellare et al's EwH.

A more favorable, efficiently searchable encryption (ESE) with PRIV security is EaH. EaH aims to model the practical scenario in database security, where a DE of some keywords works as a tag attached to the encrypted data. To search the target data, it is only required to compute the deterministic tag and compare it within the database, achieving a search time which is logarithmic in database size.

Construction of EaH. The description of the McEliece PKE is similar to the above. EaH scheme is described in Table 3. The HH property is employed in order to achieve PRIV secure efficiently searchable encryption.

[2] In this work, we do not claim any particular secure parameters. Investigating the parameters of the Hidden Hash function is out of scope of this work.

Table 2: Construction of EwH deterministic encryption

| $\mathcal{K}(1^k)$: | $\mathcal{E}(\mathsf{pk}', x)$: | $\mathcal{D}(\mathsf{sk}, \mathsf{pk}', c)$: |
|---|---|---|
| $(\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}_M(1^k)$ | Parse $\mathsf{pk}'$ into $(\mathsf{pk}, H_M)$ | Parse $\mathsf{pk}'$ into $(\mathsf{pk}, H_M)$ |
| $H_M \leftarrow \mathcal{H}(1^k, \mathsf{pk})$ | $R \leftarrow H_M(x)$ | $x, r', e \leftarrow \mathcal{D}_M(\mathsf{sk}, c)$ |
| $\mathsf{pk}' \leftarrow (\mathsf{pk}, H_M)$ | Parse $R$ to $r \Vert r_e$ | Decode $e$ to $r'_e$ |
| Return $(\mathsf{pk}', \mathsf{sk})$ | Encode $r_e$ to $e$ | $R' \leftarrow r' \Vert r'_e$ |
| | $c \leftarrow \mathcal{E}_M(\mathsf{pk}, r \Vert x; e)$ | $R \leftarrow H_M(x)$ |
| | Return $c$ | Return $x$ if $R = R'$ |
| | | Else return $\perp$ |

Proof Sketch. Our proposals derived from code-based PKE can be proven PRIV secure, since they are essentially the same as the Encrypt-with-Hash, and Encrypt-and-Hash constructions of [2]. The new and different technique we employed here is to derive the hash function (RO) from the public key itself, in a quite natural way [1, 8, 13]. Suppose that the hidden hash function in public key is a random oracle, and notice that the underlying McEliece PKE is IND-CPA secure, then it is obvious to see that the following schemes are PRIV secure, according to [2].

# 5 Deterministic Encryption from Witness-recovering PKE

A PKE which is witness-recovering encryption, decodes from the ciphertext not only the message, but also the random coin (witness) which is used to generate the ciphertext, i.e. for all $\mathcal{E}(\mathsf{pk}, x; r) = c$, $\langle x, r \rangle = \mathcal{D}(\mathsf{sk}, c)$. Code-based PKE, such as the McEliece PKE, enjoys such a specific property. Recently, there are several witness-recovering PKE [12, 3] proposed, as well. In this section, we show that since the random coin is decoded at the same time as the message, it is possible to build a high message rate DE, by using the random coin to carry some additional information. Message rate is measured as the ratio of plaintext length to ciphertext length, which characterizes the transmission efficiency. For length-preserving encryptions, such as RSA-DOAEP or our proposal in Section 3, where input length equals to output length, the message rate is optimally 1.

Although it seems somewhat strange to use randomness to carry useful information, our proposal manages to modify the random coin used in the encryption algorithm, to get higher message rate than in the original scheme, while keeping the PRIV security.

In Table. 4, let $\Lambda = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be IND-CPA secure, witness-recovering PKE scheme, where the message domain and the random coin space of $\mathcal{E}$ is $\mathcal{M}$ and $\Omega$, respectively, such that $|\mathcal{M}| = v$, $|\Omega| = \mu$. Then $\Lambda_w = (\mathcal{K}_w, \mathcal{E}_w, \mathcal{D}_w)$ is a PRIV secure DE with higher message rate.

In the above, $\mathcal{H} : \{0,1\}^* \mapsto \{0,1\}^\mu$ is considered as a family of cryptographic hash functions, i.e. random oracles. It is obvious to see that the message rate of $\Lambda_w$ is higher than before, i.e. $(|\bar{x}| + |\underline{x}|)/|c| > |\bar{x}|/|c|$, by using our new technique.

For the same reason as discussed in Section 3.1, we require that $\bar{x}$ has high min-entropy.

The security proof follows from the following two facts:

1) The basic scheme $\Lambda$ is IND-CPA secure and witness-recovering;

2) The hash function is modeled as a random oracle whose output is distributed uniformly at random.

## 5.1 Security Proof

**Theorem 2.** *Let $\Lambda = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be IND-CPA secure, witness-recovering PKE scheme. If there exists an adversary $\mathcal{A}_w$ who breaks the PRIV security of $\Lambda_w = (\mathcal{K}_w, \mathcal{E}_w, \mathcal{D}_w)$ in Table 4, then there exists an adversary $\mathcal{B}$ who breaks IND-CPA security of $\Lambda$, where the advantage of $\mathcal{B}$ is,*

$$\mathbf{Adv}^{priv}_{\Lambda_w, \mathcal{A}_w}(k) \leq \mathbf{Adv}^{ind-cpa}_{\Lambda, \mathcal{B}}(k) + 2q_h v/2^\mu + P_{\mathsf{pk}} \cdot (8q_h v + 2q_h),$$

*where $q_h$, $v$, $\mu$, and $P_{\mathsf{pk}}$ are defined in Lemma 1 below.*

*Proof.* It is concluded from Lemma 1 and the games that follow it. □

**Lemma 1.** *(Theorem 5.1 [2]) Suppose that there exists an adversary who can break the Encrypt-with-Hash (EwH) PRIV scheme with min-entropy $\mu$, which outputs vectors of size $v$ with components of length $n$ and makes at most $q_h$ queries to its hash oracle. Then there exists an IND-CPA adversary $\mathcal{B}$ against $\Lambda$ such that,*

$$\mathbf{Adv}^{priv}_{\Lambda_{EwH}, \mathcal{A}}(k) \leq \mathbf{Adv}^{ind-cpa}_{\Lambda, \mathcal{B}}(k) + 2q_h v/2^\mu + 8q_h v \cdot P_{\mathsf{pk}},$$

*where $P_{\mathsf{pk}}$ is the (maximum) probability that a public key $\mathsf{pk}$ is drawn uniformly at random from its space.*

Let us denote by $G_0, G_1, G_2$ a series of games.

Game $G_0$ The original game for PRIV security of $\Lambda_w$.

Game $G_1$ The second game is modified from $G_0$, with the only difference that we make use of a recoverable random coin $R'$ instead of $R$ in Encrypt-with-Hash $\Lambda_{EwH}$ [2], such that $R = H(\mathsf{pk} \Vert x)$ and $R' = H(\mathsf{pk} \Vert \bar{x}) \oplus \underline{x}$, where $x = \bar{x} \Vert \underline{x}$.

Table 3: Construction of EaH efficiently searchable encryption

| $\mathcal{K}(1^k)$: | $\mathcal{E}(\mathsf{pk}', x)$: | $\mathcal{D}(\mathsf{sk}, \mathsf{pk}', c\Vert T)$: |
|---|---|---|
| $(\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}_M(1^k)$ | Parse $\mathsf{pk}'$ into $(\mathsf{pk}, H_N)$ | Parse $\mathsf{pk}'$ into $(\mathsf{pk}, H_N)$ |
| $H_N \leftarrow \mathcal{H}(1^k, \mathsf{pk})$ | $T \leftarrow H_N(x)$ | $x, r, e \leftarrow \mathcal{D}_M(\mathsf{sk}, c)$ |
| $\mathsf{pk}' \leftarrow (\mathsf{pk}, H_N)$ | $r \leftarrow_R \{0,1\}^{l_p}$ | $T' \leftarrow H_N(x)$ |
| Return $(\mathsf{pk}', \mathsf{sk})$ | $e \leftarrow_R \{0,1\}^n$ | Return $x$ if $T = T'$ |
| | s.t. $Hw(e) = t$ | Else return $\perp$ |
| | $c \leftarrow \mathcal{E}_M(\mathsf{pk}, r\Vert x; e)$ | |
| | Return $c\Vert T$ | |

Table 4: Construction of deterministic encryption from witness-recovering PKE

| $\mathcal{K}_w(1^k)$: | $\mathcal{E}_w(\mathsf{pk}_w, x)$: | $\mathcal{D}(\mathsf{sk}_w, \mathsf{pk}_w, c)$: |
|---|---|---|
| $(\mathsf{pk}, \mathsf{sk}) \leftarrow_R \mathcal{K}(1^k)$ | Parse $\mathsf{pk}_w$ to $(\mathsf{pk}, H)$ | Parse $\mathsf{pk}_w$ to $(\mathsf{pk}, H)$ |
| $\mathcal{H} : \{0,1\}^* \mapsto \{0,1\}^\mu$ | Parse $x$ to $\bar{x}$, $\underline{x}$, | $\mathsf{sk} \leftarrow \mathsf{sk}_w$ |
| $H \leftarrow_R \mathcal{H}(1^k)$ | s.t. $\lvert\bar{x}\rvert = v$, $\lvert\underline{x}\rvert = \mu$ | $\langle \bar{x}, R\rangle \leftarrow \mathcal{D}(\mathsf{sk}, c)$ |
| $\mathsf{pk}_w \leftarrow (\mathsf{pk}, H)$ | $R \leftarrow H(\mathsf{pk}\Vert\bar{x}) \oplus \underline{x}$ | $\underline{x} \leftarrow H(\mathsf{pk}\Vert\bar{x}) \oplus R$ |
| $\mathsf{sk}_w \leftarrow \mathsf{sk}$ | $c \leftarrow \mathcal{E}(\mathsf{pk}, \bar{x}; R)$ | $x \leftarrow \bar{x}\Vert\underline{x}$ |
| Return $(\mathsf{pk}_w, \mathsf{sk}_w)$ | Return $c$ | Return $x$ |

Game $G_2$ The third game is modified from $G_1$, with the only difference that $\mathcal{A}_f$ have queried $H(\mathsf{pk}||\bar{x}) \oplus \underline{x}$ before $\mathcal{A}_g$ is initiated.

We borrow the proof of Encrypt-with-Hash scheme from [2], which only differs from ours in that we make use of a recoverable random coin $R'$ instead of $R$ in [2], s.t. $R = H(\mathsf{pk}||x)$ and $R' = H(\mathsf{pk}||\bar{x}) \oplus \underline{x}$, where $x = \bar{x}||\underline{x}$. Thus, it only needs to be shown that the distributions of $R$ and $R'$ are uniform and random, so that no adversary can distinguish them. Thanks to employment of the random oracle $H$, we can simply reuse the random coin to carry the message as well as to build the proof.

Thus, we have $|\Pr[G_1] - \Pr[G_0]| \leq \mathcal{A}_{RO}$. Since we have assumed the random oracle model, then $\mathcal{A}_{RO}$ is zero and $\Pr[G_1] = \Pr[G_0]$.

On the other hand, for the game $G_2$ obtained from $G_1$, it is crucial to observe that any adversary $\mathcal{A}_f$ may have queried $H(\mathsf{pk}||\bar{x}) \oplus \underline{x}$ before $\mathcal{A}_g$, so that the simulation might fail. However, this probability is bounded by $2P_{\mathsf{pk}} \cdot q_h$.

Thus, there is $|\Pr[G_2] - \Pr[G_1]| \leq 2P_{\mathsf{pk}} \cdot q_h$. The factor of 2 comes from the fact that $\mathcal{A}_f$ has control over two distinct messages. As a consequence, the probability of simulation failure is negligible as long as $P_{\mathsf{pk}}$ is negligible. Note that the later property is not the standard requirement for PKE, however it holds for all known PKE.

Notice that $G_2$ is the PRIV security game of $\Lambda_{EwH}$, we have the following,

$$\mathbf{Adv}^{priv}_{\Lambda_w, \mathcal{A}_w}(k) \leq \mathbf{Adv}^{priv}_{\Lambda_{EwH}, \mathcal{A}}(k) + 2P_{\mathsf{pk}} \cdot q_h.$$

Summarizing the above and Lemma 1, we have finished the proof.

# 6 Extension to Chosen-Ciphertext Security

Above, we have proposed several PRIV secure DE schemes in the CPA case. We believe that it is possible to extend our results to the CCA scenario. As commented in [2], a PRIV-CCA scheme could be obtained from a PRIV-CPA one with some additional cost, such as one-time signatures or other authenticated techniques to deny a decryption query from the CCA attacker. The important point is that we have achieved very efficient PRIV-CPA secure building blocks, which in some aspects are better than the previously known ones. A bad news is that when extending RSA-DOAEP based hybrid encryption to the CCA scenario, it will probably lose its nice length-preserving property, because some consistency check raises the overhead of bandwidth.

# 7 Conclusion

In the random oracle model, we presented a generic and efficient construction of deterministic hybrid encryption by composing PRIV-secure PKE and IND-CPA SKE. In particular, this construction implies the first length-preserving DHE, when instantiated with RSA-DOAEP.

Moreover, we presented a postquantum deterministic encryption by plugging-in the McEliece PKE into the generic constructions Encrypt-and-Hash and Encrypt-with-Hash by Bellare and Boldyreva [2]. We point out that the McEliece public key can also be used as a hash function.

Furthermore, we showed that witness-recovering encryption can be used to construct deterministic encryption schemes with plaintext length which is larger than that of the original schemes, by using a part of the plaintext as random coins (witness).

Finally, we noted that the standard authentication techniques (discussed, in particularly, in [2]) can be used to upgrade our schemes to PRIV-CCA security.

All the above results come for the price of assuming a particular distribution of the plaintext – namely that its first part (having a size of the domain of the underlying PRIV-secure PKE scheme) is of high min-entropy. An open question is to replace this assumption with a standard one – the high min-entropy of the whole plaintext. Furthermore, since all the above results are achieved in the random oracle model, another open question is to remove this assumption by obtaining the constructions secure in the standard model.

# Acknowledgments

The authors would like to thank the anonymous reviewer of IEICE Transactions of Fundamentals of Electronics, Communications and Computer Sciences for pointing out a trivial attack (described in Section 3) in an earlier version of this work. The first author would like to thank the support by Start-up Grant-in-Aid for Young Scientists, Japan Society for the Promotion of Science (JSPS), No. 21800094.

# References

[1] D. Augot, M. Finiasz, and N. Sendrier, "A family of fast syndrome based cryptographic hash functions," *Mycrypt '05*, vol. LNCS 2715, pp. 64–83, 2005.

[2] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," *Crypto '07*, vol. LNCS 4622, pp. 535–552, 2007.

[3] M. Bellare, M. Fischlin, A. O'Neill, and T. Ristenpart, "Deterministic encryption: Definitional equivalences and constructions without random oracles," *Crypto '08*, vol. LNCS 5157, pp. 360–378, 2008.

[4] M. Bellare and P. Rogaway, "Random oracles are practical: A paradigm for designing efficient protocols," in *ACM Conference on Computer and Communications Security*, pp. 62–73, 1993.

[5] A. Boldyreva, S. Fehr, and A. O'Neill, "On notions of security for deterministic encryption, and efficient constructions without random oracles," *Crypto '08*, vol. LNCS 5157, pp. 335–359, 2008.

[6] T. Cover, "Enumerative source encoding," *IEEE IT*, vol. 19, no. 1, pp. 73–77, 1973.

[7] R. Cramer and V. Shoup, "Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack," *SIAM Journal on Computing*, vol. 33, no. 1, pp. 167–226, 2003.

[8] M. Finiasz, "Syndrome based collision resistant hashing," *PQCrypto'08*, vol. LNCS 5299, pp. 137–147, 2008.

[9] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270–299, 1984.

[10] R. J. McEliece. "A public-key cryptosystem based on algebraic coding theory,". Tech. Rep. Deep Space Network Progress Report 42-44, 1978.

[11] R. Nojima, H. Imai, K. Kobara, and K. Morozov, "Semantic security for the mceliece cryptosystem without random oracles," *Designs, Codes and Cryptography*, vol. 49, no. 1-3, pp. 289–305, 2008.

[12] C. Peikert and B. Waters, "Lossy trapdoor functions and their applications," in *STOC 2008*, pp. 187–196, 2008.

[13] J. Stern, "A new identification scheme based on syndrome decoding," *Crypto '93*, vol. LNCS 773, pp. 13–21, 1993.

**Yang Cui** received his B.E. degree in electronic and communication engineering from Harbin Institute of Technology, China, in 2000, and M.E., Ph.D. degree in information and communication engineering from the University of Tokyo, Japan, in 2004, 2007 respectively. From 2007 to 2010, he was enrolled with Research Center for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST). Since 2011, he is with Wireless Network Research Department, Huawei Technologies Co. Ltd., China. His research interests include cryptology, network security, telecommunication security and coding theory.

**Kirill Morozov** received his M.E. degree in radio engineering from Saint-Petersburg State University of Telecommunications, Russia, in 1998 and Ph.D. degree in computer science from the University of Aarhus, Denmark, in 2005. In 2005, he was a postdoctoral fellow at the University of Tokyo. He has been a research scientist at the National Institute of Advanced Industrial Science and Technology (AIST) in 2006-2010. From December 2010 to March 2011 he was an assistant professor in the Faculty of Mathematics, Kyushu University. From April 2011, he is an assistant professor in the Institute of Mathematics for Industry, Kyushu University. His research interests include cryptography and information security. He is a member of IEICE and IACR.

**Kazukuni Kobara** received his Ph.D. degree in engineering from the University of Tokyo in 2003. From 2000 to 2006 he was a research associate at the Institute of Industrial Science of the University of Tokyo. In 2006, he joined the National Institute of Advanced Industrial Science and Technology (AIST) where he was the leader of the Research Team for Security Fundamentals in the Research Center for Information Security (RCIS). Currently he is the leader of the Control System Security Research Group of the Research Institute for Secure

Systems (RISEC) as well as CTO of AIST Technology Transfer Venture, BURSEC Inc. His research interests include cryptography, information and computer security. He received the SCIS Paper Award and the Vigentennial Award from IEICE in 1996 and 2003, respectively. He also received the Best Paper Award of WISA, the ISITA Paper Award for Young Researchers, the IEICE Best Paper Award and Inose Award, the WPMC Best Paper Award and the JSSM Best Paper Award in 2001, 2002, 2003, 2005 and 2006 respectively. He is a member of IEICE and IACR (International Association for Cryptologic Research). He served as a member of CRYPTREC (Cryptography Research and Evaluation Committees) in Japan from 2000 to 2008, the vice chairperson of MIC WLAN security committee in 2003 and the chief investigator of INSTAC identity management committee from 2007-2009.

**Hideki Imai** received the B.E., M.E., and Ph.D. degrees in electrical engineering from the University of Tokyo in 1966, 1968, and 1971, respectively. From 1971 to 1992 he was on the faculty of Yokohama National University. From 1992 to 2006 he was a Professor in the Institute of Industrial Science, the University of Tokyo. In 2006 he was appointed as an Emeritus Professor of the University of Tokyo and a Professor of Chuo University. Concurrently he served as the Director of Research Center for Information Security, National Institute of Advanced Industrial Science and Technology from 2005 to 2011. From IEICE he received Best Book Awards in 1976 and 1991, Best Paper Awards in 1992, 2003, 2004 and 2008, Yonezawa Memorial Paper Award in 1992, Achievement Award in 1995, Inose Award in 2003, and Distinguished Achievement and Contributions Award in 2004. He also received Golden Jubilee Paper Award from the IEEE Information Theory Society in 1998, Wilkes Award from the British Computer Society in 2007, and Official Commendations from the Minster of Internal Affairs and Communications in 2002, from the Minister of Economy, Trade and Industry in 2002, and from the Chief Cabinet Secretary in 2009. He was awarded Honor Doctor Degree by Soonchunhyang University in 1999 and Docteur Honoris Causa by Toulon University in 2002. He is also the recipient of the Ericsson Telecommunications Award 2005 and the Okawa Prize 2008. Dr. Imai is a member of the Science Council of Japan. He was elected a Fellow of IEEE, IEICE, and IACR in 1992, 2001, and 2007, respectively. He is now an IEEE Life Fellow and an IEICE Fellow, Honorary Member. He served as the President of the Society of Information Theory and its Applications in 1997, of the IEICE Engineering Sciences Society in 1998, and of the IEEE Information Theory Society in 2004. He was the Chair of the IEEE Tokyo Section from 2009 to 2010. He is currently the Chair of CRYPTREC (Cryptography Techniques Research and Evaluation Committee of Japan).