Longer version of an extended abstract to be published in Advances in Cryptology—ASIACRYPT 2007, Springer-Verlag, 2007. Available online from: `http://www.cs.stanford.edu/~xb/asiacrypt07/`

Miniature CCA2 PK Encryption : Tight Security Without Redundancy

Xavier Boyen
Voltage Inc. — *xb@boyen.org*

November 12, 2007

Abstract

We present a minimalist public-key cryptosystem, as compact as ElGamal, but with adaptive chosen-ciphertext security under the gap Diffie-Hellman assumption in the random oracle model. The novelty is a dual-hash device that provides tight redundancy-free implicit validation. Compared to previous constructions, ours features a tight security reduction, *both in efficacy and efficiency*, to a classic and essentially non-interactive complexity assumption, and without resorting to asymmetric/symmetric-key hybrid constructions. The system is very compact: on elliptic curves with 80-bit security, a 160-bit plaintext becomes a 320-bit ciphertext. It is also very simple and has a number of practical advantages, and we hope to see it adopted widely.

1 Introduction and Motivation

One of the major pursuits in cryptographic research has been to devise faster, nimbler, shorter, and stronger encryption systems that can be used in practice. In the realm of public-key cryptosystems, the lure of simplicity and efficiency has produced many a breakthrough and many more successive refinements, over the last three decades.

We propose one more such technical refinement, in the form of a CCA2-secure PK cryptosystem with the shortest ciphertext among Discrete-Log-based systems at any given *exact* security level. Our construction is simple and purely algebraic, and relies on a standard assumption in the random-oracle model of [2]. To obtain short ciphertexts, we eliminate all sources of redundancy, and limit the unavoidable randomness to a single element of the computational group. Furthermore, we ensure that no space is wasted in the encoding of that element, by shrinking the computational group itself to the smallest size that the birthday paradox will allow. The latter requirement is only possible with a *tight* reduction to the underlying security assumption, as we shall discuss momentarily. These properties taken together account for the scheme's compactness.

All comparable schemes that have been suggested over the years either have a non-tight security reduction, or are hybrid constructions with both an algebraic and a symmetric-key component, each bringing forth its own complexity assumption. (We note that all known redundancy-free systems depend either on some non-standard oracle assumption, or at least on the random-oracle model; indeed, it remains an open problem to withstand active attacks without redundancy and without relying on random oracles or some sort of interactive assumption.)

By contrast, the scheme we propose can be proven tightly secure, in the random-oracle model, solely under the Gap Diffie-Hellman (Gap-DH) assumption [25], or even under the usual Computational Diffie-Hellman (CDH) assumption if the algebraic group admits an efficient bilinear pairing:

This is because with a pairing one can instantiate the DDH oracle posited by the Gap-DH assumption, which then reduces to plain CDH. Pairing-friendly groups are easy to construct on certain types of elliptic curves; we refer the reader to the abundant literature on pairing-based cryptography. We emphasize that our scheme will be secure under CDH as soon as a pairing *exists* in the selected group, even though we never actually *use* it. In groups where no efficient pairing is known to exist, security still follows from the Gap-DH assumption.

1.1 On the Tightness of Reductions

The importance of a tight security reduction to a simple and well-studied assumption is crucial to the determination of the exact security of any cryptosystem. A security proof can be loose in two different ways: the final reduction may cause a loss of success probability, or the simulator can be slow and steal most of the computational time that should go to the attacker. The latter factor is too often ignored when a security proof is advertised as tight: it is often the case that a proof with tight efficacy probability-wise, would use an inefficient simulator whose running time is quadratic or worse, which can significantly hurt the security of the scheme in a real-world attack: the true security guarantee would *not* be tight if one accounted for all parameters, as one should. Accordingly, it is only by taking into account all intervening factors that a scheme's *true security* can be determined for a chosen *apparent security parameter*. Larger apparent security parameters will have to be selected to compensate for loose reductions (or strong assumptions), resulting in larger ciphertexts for the desired target security level.

In the random oracle model in particular, it is very important to pay close attention to the extent that a scheme's exact security deteriorates with the number of random oracle queries made by the adversary, because in reality the random oracle is instantiated as an algorithmic hash function that can be queried offline, limited only by the adversary's computational powers. Interactive assumptions that assume the existence of "fancy oracles" that have no actual instantiations (even imperfect ones) are even more troublesome, because there is no telling how a scheme that depends on such an oracle will fare in the real world: it might be completely insecure and it is not hard to find examples of such. Sensitivity to the number of decryption queries is less critical because in practice decryption queries are implemented as active attacks with naturally slower query rates, but it remains an issue nevertheless.

1.2 Our Contribution

For all of the reasons above, it is our purpose here to devise a compact encryption scheme based on plausible assumptions, and establish exact security bounds in function of the number of random-oracle and decryption queries made by the opponent. We seek to obtain a tight security bound that is quasi-independent of the number of those queries (as long as their number remains sub-exponential in the security parameter, which is an unavoidable requirement). Surely, eliminating the random oracle itself would be even more desirable, but it is an open problem whether that is even feasible at all if no redundancy whatsoever is tolerated.

Our main technical tool stems from the observation that a pair of sequential one-time pads, can, in the random-oracle model, give us an almost tight reduction from a mild assumption such as CDH or Gap-DH, without appealing to explicit ciphertext redundancy or a hybrid scheme. Whereas redundancy-free public-key schemes with a tight reduction have been proposed in the past, we view the dual-hash device and the simpler structure that it enables as our main contributions. As an added bonus, our system will support very efficient non-interactive threshold decryption.

2 Toward Active Security Without Redundancy

The most common threat to CCA2 security is that of a query on a malformed ciphertext causing the decryption oracle to leak damaging information, either about the private key, or about the plaintext (when the malformed ciphertext is a deformation of a legitimate one). For this reason, the most common way to construct a CCA-secure system from a CPA-secure one is to add some redundancy, thanks to which malformed or mauled ciphertexts can be safely rejected. Redundancy has also an utilitarian purpose in the security proofs: simulators use it to extract private knowledge about the ciphertext creation, which gives them a backdoor thanks to which decryption queries can be answered without knowledge of the decryption key. The two main ways that this backdoor is implemented are the NIZK and IBE approaches, briefly described below.

Redundancy can nevertheless be avoided provided that the decryption of malformed ciphertexts is made harmless, *e.g.*, as will be the case if the decryption of bogus ciphertexts appears uniformly random to the adversary. Thus, as has been observed several times before, redundancy is not truly necessary in order to achieve chosen-ciphertext security (though randomness is always needed for semantic security). Technically, one must also ensure that the simulator is still able to answer the decryption queries in the absence of a redundancy backdoor: this is where idealized models such as the random oracle heuristic [2] must come into play, at least in our current state of knowledge.

Subject to the above limitations, there exists a rather extensive body of work on public-key encryption systems secure against active attacks. We now review the main proposals, concentrating on systems that are usable in practice. In order to depict a more complete landscape, we also discuss a number of redundant constructions, since they far outnumber the redundancy-free ones. Once again, if we insist on the lack of redundancy, no CCA2-secure public-key systems, not even conjectured ones, are known to exist in the standard model.

2.1 In the Standard Model

In the standard model, all known chosen-ciphertext-secure systems require some redundancy.

First of all, we mention the early theoretical work of Dolev, Dwork, and Naor [17], which achieves CCA-security using a bitwise construction which is too inefficient to be used in practice. More efficient constructions were to follow, based either on the so-called two-key paradigm, or, more recently, on identity-based encryption and related techniques.

The two-key (or double-encryption) framework for chosen-ciphertext security was first proposed by Naor and Yung [35], and perfected by Cramer and Shoup [14] who gave the first efficient CCA2-secure public-key scheme in the standard model. There were many subsequent improvements to the Cramer-Shoup system, and the current state of the art is due to Kurosawa and Desmedt [30]. The two-key paradigm consists in providing two independent encryptions of the same plaintext, along with a Non-Interactive Zero-Knowledge (NIZK) proof that the two plaintexts are the same. This provides the needed redundancy that allows the simulator to answer decryption queries. A drawback of this approach is that the redundancy cannot be checked until the complete ciphertext has been decrypted, which makes threshold decryption a complicated proposition [11, 22, 38].

The Identity-Based Encryption (IBE) approach was recently proposed by Canetti, Halevi, and Katz [12], and subsequently improved [8, 10]. Here, the general idea is to encrypt a plaintext to an identity equal to a signature verification key, or some function of the ciphertext itself, that the recipient can use to authenticate the ciphertext. This is a different kind of redundancy that leads to a completely different type of simulation proof than in the two-key approach. Both methods are comparable in terms of efficiency. One advantage of the identity-based approach is that the integrity check can be done before decryption, which makes non-interactive threshold decryption easy [6]. The main disadvantage of the IBE approach is that it uses bilinear pairings, although it is possible to eliminate them entirely by making stronger assumptions [27, 28].

Although reasonably efficient, all these constructions require at least two group elements' worth of ciphertext overhead. It is an open problem to achieve chosen-ciphertext security without redundancy in the standard model.

2.2 Using Random Oracles

In parallel to the above developments, researchers have sought to construct CCA2-secure systems with efficiency as the primary goal, even if that meant using the random oracle heuristic. One of the most significant works in this area is the RSA-OAEP padding scheme [3] and its subsequent improvements [4, 37], which are widely deployed as a standard. However, the development of OAEP was tormented: the original redundancy-free design had to be scrapped in order to achieve provable chosen-ciphertext security, and it took several years until an RSA system with both properties was finally invented (see below).

In parallel, a powerful result by Fujisaki and Okamoto [19], subsequently improved by the same authors [20], shows that any CPA-secure encryption scheme can be generically transformed into a CCA2-secure one, in the random oracle model, simply by adding some judicious redundancy. One can thus assemble a very efficient CCA2-secure system simply by taking an elliptic-curve implementation of the ElGamal cryptosytem and applying the Fujisaki-Okamoto transformation. This does however introduce some redundancy.

2.3 From Interactive Hash Assumptions

Since random oracles alone did not seem sufficient to obtain redundancy-free chosen-ciphertext security, one had to appeal to more exotic and stronger assumptions. In general, these assumptions are interactive and involve at least a random function, very much like the random oracle model.

The first system to achieve redundancy-free chosen-ciphertext security, is that of Phan and Pointcheval [36]. The Phan-Pointcheval scheme can be thought of as an extension of RSA-OAEP that achieves adaptive security using the theoretical minimum amount of randomness and no redundancy, but under a strong non-standard interactive assumption. Roughly speaking, it combines a trapdoor permutation with an idealized random permutation; the CCA2 security proof then holds in the random permutation model. In practice, the system is instantiated using RSA and a Feistel network, which only requires a random oracle rather than a random permutation.

The second system in this category is DHIES [1], all of whose variants are based on a strong interactive assumption known as Oracle Diffie-Hellman. The DHIES system is a hybrid of ElGamal, a symmetric cipher, and a MAC, and is provably secure under the ODH assumption. Because of the MAC, the original DHIES system is not redundancy-free.

Kurosawa and Matsuo [31] subsequently gave an improvement to DHIES that eliminated the MAC from the ciphertext and thus the redundancy. This was done by means of a special "all-or-nothing" mode of operation for the symmetric cipher, such as CMC [23] and EME [24], which can be viewed as an analog to the pseudo-random permutation in the Phan-Pointcheval system. With this modification, DHIES no longer incurs any expansion, and thus the Kurosawa-Matsuo system is indeed free of redundancy. Since furthermore DHIES can be implemented on elliptic curves, unlike

Phan-Pointcheval which uses integer arithmetic modulo a large RSA composite, Kurosawa-Matsuo can be made very compact. Indeed, their system currently holds the record for the most compact CCA2 public-key system for short messages.

Libert and Quisquater [32] later transposed the ideas of Kurosawa and Matsuo to the identity-based encryption setting, and in particular to the IBE system of Boneh and Franklin [7]. They show that CCA2 security can be obtained by using an expansion-less chosen-ciphertext-secure symmetric mode of operation (instead of the Fujisaki-Okamoto transformation as originally used in [7]). The Libert-Quisquater IBE system is in fact simpler than the Kurosawa-Matsuo PKE, but unfortunately, the security of the former rests (in the RO model) upon a very strong interactive assumption called Gap Bilinear Diffie-Hellman, which is not even falsifiable in our current state of knowledge since nobody knows how to construct a Gap-BDH challenger.

To conclude this tour, we now briefly review the main features of the Phan-Pointcheval and the Kurosawa-Matsuo systems, as these are the two schemes against our construction ought to be compared.

2.3.1 The Phan-Pointcheval System

Phan and Pointcheval [36] gave the first construction of a CCA2-secure public-key encryption system without redundancy. It is based on the RSA trapdoor permutation which is made non-malleable using a idealized random permutation instantiated as a Feistel network. The Phan-Pointcheval system incurs very little ciphertext expansion: for an *apparent* security parameter κ, the ciphertext is only κ bits longer than the message it encrypts. Without taking the security reduction efficiency into account, this is the smallest possible ciphertext expansion that can be achieved by any public-key encryption scheme at the $2^{-\kappa}$ security level.

In reality, Phan-Pointcheval is not quite as compact as we would like, for a couple of reasons: (1) its security reduction has tight efficacy but only quadratic efficiency in the Feistel network instantiation, which means that in practice its exact security could degrade significantly with the number of queries made by the adversary, which ought to be compensated by growing the modulus; (2) because the scheme is built around an RSA permutation, ciphertexts cannot be made smaller than 1024 bits at the 2^{-80} security level, or 15360 bits at the 2^{-256} security level, to guard against sub-exponential factorization attacks of complexity $L(1/3)$ using the number field sieve.

2.3.2 The Kurosawa-Matsuo System

To avoid the minimum size limitation associated with RSA groups, Kurosawa and Matsuo [31] have proposed a different construction of a CCA2-secure public-key cryptosystem, based not on RSA but on ElGamal. Since ElGamal can be implemented on elliptic curves, much fewer bits are in principle needed in order to achieve the same security. The Kurosawa-Matsuo construction is set in the KEM/DEM framework, where a CCA2-secure KEM is constructed simply by hashing an ElGamal session key, from which an expansion-less one-time chosen-ciphertext-secure DEM is used to encrypt the actual message. For an *apparent* security parameter κ, the ciphertext is 2κ bits longer than the message, which is the smallest possible expansion for a Discrete-Log-based cryptosystem, due to the birthday bound barrier associated with generic discrete-log attacks.

On the negative side, the security reduction of the Kurosawa-Matsuo system relies on the original DHIES construction, which is based on a rather strong interactive assumption called the Oracle Diffie-Hellman assumption. Roughly speaking, the ODH problem asks us to distinguish

(g, g^a, g^b, g^{ab}) from (g, g^a, g^b, g^r) given access to an oracle $\mathcal{O} : h \mapsto H(h^a)$, which can be thought of as the composition of the composition of a secret-power exponentiation with an ideal random hash function (also kept secret by default). We note however that Cramer and Shoup [16] later gave an alternative security proof of DHIES, replacing ODH with Gap-DH in the random oracle model. Their proof should also apply to the Kurosawa-Matsuo system.

Perhaps the main downside of the Kurosawa-Matsuo system is that it depends on rather complex modes of operation for block ciphers, such as the deterministic, redundancy-free, one-time chosen-ciphertext-secure modes given in [23, 24]. Because of those extraneous components, the Kurosawa-Matsuo system may suffer from a larger implementation footprint than competing schemes. The complex modes of operation may also pose practical challenges for arbitrary-size plaintexts.

2.4 The New Construction

Here, we propose another efficient public-key encryption system without redundancy and with a tight adaptive chosen-ciphertext security proof. A feature of our scheme is its simple and self-contained algebraic structure. The security reduction is to the Gap Diffie-Hellman assumption in the random-oracle model. Gap-DH is a "decisional/computational gap" assumption [25], which simply posits that CDH is hard given a DDH oracle. Since Gap-DH itself reduces to the usual CDH in groups equipped with a bilinear map (which we know how to construct), our scheme belongs with the "plain" random-oracle schemes of Section 2.2, as opposed to the "fancy" interactive-assumption schemes of Section 2.3, which until now were the only ones known to avoid redundancy. Practically speaking, our system only uses hashing and generic group arithmetic (no block cipher and no complex mode of operation), and so its implementation should be straightforward in any programming language with a decent library.

The main idea of the scheme is to blind the message not once, but twice, using ElGamal one-time pads that are homomorphically related to the same secret decryption key. The resulting ciphertext has no explicit redundancy because the second key can be reconstructed from the first without having to include any information about it. In the random oracle model, this however gives us the implicit consistency check needed for chosen-ciphertext security. Furthermore we can simulate it in constant time and almost perfectly (*i.e.*, with negligible security loss) against any polynomially bounded adversary, hence the tight security.

2.4.1 Security and Compacity

It should be mentioned that it does not seem feasible to achieve a better "ciphertext compacity *vs.* exact security" tradeoff without leaving the realm of Discrete-Log-based algebraic CCA2 PKE systems. Indeed, at the $2^{-\kappa}$ exact security level, the ciphertext overhead is a single group element, which takes as few as 2κ bits to represent; however, the randomness embedded in this element cannot be removed, and any attempt to reduce the entropy of that group element further will enable a generic discrete logarithm attack of relative complexity lower than $\sqrt{2^{2\kappa}} = 2^{\kappa}$.

However, one should not infer from this that shorter ciphertexts are not possible using different techniques. For example, with trapdoor permutations it is possible to reduce the overhead to the theoretical minimum of κ bits, as in the Phan-Pointcheval system; one problem with this approach is that RSA-based trapdoor permutations require much larger groups than elliptic curves for the same security (which is why Phan-Pointcheval ciphertexts remain large despite the very low overhead). Substituting a more compact trapdoor permutation for RSA in Phan-Pointcheval would be an excellent way to create a more compact scheme than the present proposal. Of course, constructing a compact trapdoor permutation in the first place, *e.g.*, whose inputs and outpus are no greater than 3κ bits at the $2^{-\kappa}$ security level, is another long-standing famous open problem in cryptography.

2.4.2 State of the Art

We do not claim that our construction constitutes a deep result, but merely a practical one that we hope will be adopted in practice. In retrospect, our construction and its security proof appear quite simple, indeed, as surely many other results of this sort have before it. However, the fact that with a simple trick we have improved upon the state of the art on an old problem is a compelling indication that there are still new insights to be gained in this area. Thus we hope that this contribution will be useful to security practitioners, and perhaps inspire new ideas to researchers in the field.

3 The Miniature CCA2 System

We are now almost ready to present the construction. Unlike Kurosawa and Matsuo, we seek to build an integrated encryption scheme without insisting on a separation between KEM and DEM. On the contrary, we look for an algebraic construction that avoids block ciphers and their complex modes of operations, and seek to base our scheme on a single mild and well-studied assumption.

3.1 Inching toward a Solution

Before we present our construction, it is useful to try out a few approaches, to see what works and what does not. This will make it easier to understand the design of the final scheme.

1. To start, consider the hashed ElGamal system, whose ciphertext is $(c_1, c_2) = (M \oplus H(g_1^r),\ g_2^r)$ for random $r \in \mathbb{F}_p$. The public key is $(g_1, g_2) \in \mathbb{G}^2$, and the decryption key is $k = \mathrm{dlog}_{g_1}(g_2)$. The ciphertext is free of redundancy, but it is malleable and thus the scheme is only secure under passive attacks.

2. To make the scheme secure under active attacks, we can modify the ciphertext as follows: $(c_1, c_2) = (M \oplus H_1(g_1^r),\ g_2^r {g_3^r}^{H_2(c_1)})$, where H_1 is viewed as a random oracle and H_2 is collision resistant. The public key is (g_1, g_2, g_3) and the secret key their discrete logs.

   Here, there is no obvious active attack, and in fact the scheme can be proven IND-CCA2 secure under the Gap-DH assumption in the random oracle model. Unfortunately, the reduction is not tight, and is in fact rather expensive because, for each decryption query, the DDH oracle must be tested against the inputs to all previous random-oracle queries.

3. The reduction in the previous scheme can be made more efficient, and thus the scheme more secure in the exact sense, by including more information inside the random-oracle input, as in: $(c_1, c_2) = (M \oplus H_1(g_1^r, g_2^r),\ g_2^r {g_3^r}^{H_2(c_1)})$. We can also take $g_3 = g_1$ to make the key shorter.

   This simple modification greatly reduces the number of DDH oracle queries needed by the simulator (in a security reduction to Gap-DH), to the point that we now have proportionality between the adversary's and the simulator's use of their respective oracles, *i.e.*, one query to the DDH oracle for each random-oracle query. The resulting reduction is thus more efficient, and, indeed, public-key systems with this exact structure have been recently and independently suggested in at least two places [33, 29], prior to the publication of this work.

However, security still is not tight. For *every* decryption query, the simulator must perform a non-trivial group operation between c_2 and the input to *every* random oracle query made so far. Thus, if the adversary makes q_d decryption and q_H random-oracle queries, the simulator's running time will be at least the product of the two, *i.e.*, $\Omega(q_d q_H)$, which is clearly disproportionate (*i.e.*, super-linear) to the sum total of all of the adversary's queries.

Hence, although the efficacy or succcess probability of the reduction may be tight, and the use of the DDH oracle parsimonious, the reduction algorithm remains inefficient due to an excess of bookeeping.

A general principle that emerges from these examples is how random oracles can be utilized to extract the information needed to answer decryption queries, when the ciphertext contains no redundancy that would let us do so in another way (as in the schemes mentioned in Section 2.1).

We can also see, in all these examples and analogous constructions based on a Gap assumption, that the simulator must try out all random oracle inputs to see if one works for every decryption query that it answers. This is not unrelated to the fact that our assumption (Gap-DH) only provides a decisional (yes/no) oracle to the simulator, and perhaps one of the reasons the Kurosawa-Matsuo scheme avoids this problem is because its DHIES component relies on a stronger assumption.

However, the central reason for the schemes' reduction inefficiency is their use of a single random oracle for blinding the message (as in $M \oplus H(...)$). It turns out that a much more efficient simulator can be made if we had two random-oracle one-time pads to play with (as in $M \oplus H_1(...) \oplus H_2(...)$). Why this is so will become apparent when we construct a simulator in Section 3.4.

Remark: KEM without DEM. Incidentally, we note that the KEM-only version of hashed ElGamal from Item 1 above, namely, with ciphertext $c = g_2^r$ and random session key $k = H(g_1^r)$, was already redundancy-free and secure against active attacks (under Gap-DH in the random-oracle model). Most of the difficulties we encountered above were caused by the addition of an improper or suboptimal DEM component.

This is another illustration that the secure encryption of a random key can sometimes be easier to achieve than that of an arbitrary message; see also [10] for a different manifestation of this phenomenon.

3.2 The Full Scheme

Our construction is based on some of the principles hinted to above. The main difficulty is to obtain a double one-time-pad blinding of the message without lengthening the ciphertext, and then to use this double blinding in the security proof to achieve a tight reduction.

We start with the construction, which uses two random oracles Φ and Ψ, and one collision-resistant function π which could be a simple injection.

Context: Let $\kappa \in \mathbb{N}$ be an arbitrary security parameter. Let $\mathbb{G} = \langle U \rangle$ be a cyclic prime-order group (written multiplicatively), generated by U, of prime order p, such that $2^{2\kappa-1} < p < 2^{2\kappa+1}$. Let $\mathbb{F}_p$ be the finite field of size p, and let $\mathbb{F}_p^\times = \mathbb{F}_p \setminus \{0\}$ denote its multiplicative group of order $p-1$. Let $\mathcal{M} = \{0,1\}^\ell$ be the set of all bit strings of length ℓ, for any fixed $\ell \geq 2\kappa$.

Let $\pi : \mathcal{M} \to \mathbb{F}_p^\times$ be an arbitrary injection or a collision-resistant hash function.
Let $\Phi : \mathbb{G} \times \mathbb{G} \to \mathcal{M}$ and $\Psi : \mathbb{G} \to \mathcal{M}$ be two cryptographic hash functions (viewed as RO).

Key generation: Draw a secret random exponent $s \in_\$ \mathbb{F}_p^\times$, and calculate $V \leftarrow U^s$.

The public encryption key is $\mathsf{Pk} \leftarrow (U, V) \in \mathbb{G}^2$.
The private decryption key is $\mathsf{Sk} \leftarrow s \in \mathbb{F}_p^\times$.

Encryption: Given Pk and a plaintext $\mathsf{Msg} \in \mathcal{M}$, pick a randomizer $r \in_\$ \mathbb{F}_p^\times$, and let,

$$
\begin{aligned}
&A \leftarrow V^r \\
&B \leftarrow \Psi(A) \oplus \mathsf{Msg} \\
&C \leftarrow V^{r/\pi(B)} \\
&D \leftarrow U^{r/\pi(B)} \\
&E \leftarrow \Phi(D, C) \oplus B
\end{aligned}
$$

The ciphertext is $\mathsf{Ctx} = (D, E) \in \mathbb{G} \times \mathcal{M}$.

Decryption: Given Sk and a ciphertext $\mathsf{Ctx} = (\bar{D}, \bar{E})$, check that $1 \neq \bar{D} \in \mathbb{G}$, and let,

$$
\begin{aligned}
&\bar{C} \leftarrow \bar{D}^{\mathsf{Sk}} \\
&\bar{B} \leftarrow \bar{E} \oplus \Phi(\bar{D}, \bar{C}) \\
&\bar{A} \leftarrow \bar{C}^{\pi(\bar{B})} \\
&\bar{M} \leftarrow \bar{B} \oplus \Psi(\bar{A})
\end{aligned}
$$

The decrypted plaintext is $\bar{\mathsf{Msg}} = \bar{M} \in \mathcal{M}$.

3.3 Operational Efficiency

Encryption and decryption have essentially the same computational costs, which are dominated by the costs of two exponentiations in $\mathbb{G}$, plus (for long messages) two passes on a buffer whose size is that of the input string (resp. plaintext or ciphertext). In particular, we note the following:

- Encryption requires only two exponentiations (and not three), because most of the work done to compute V^r can be reused to compute $V^{r/\pi(B)}$, regardless of the exponentiation algorithm used (whether straight double-and-add, or one of the many efficient window methods; *cf.*, *e.g.*, [34]).
- Decryption can similarly be performed in about a single exponentiation (instead of two), by computing $\bar{C}^{\pi(\bar{B})}$ as $\bar{D}^{\pi(\bar{B})\cdot\mathsf{Sk}}$, which uses the same generator as $\bar{D}^{\mathsf{Sk}}$ and thus shares the same intermediate powers.
- In both cases, only two passes on the buffer are needed (and not three): first on the input string Msg or $\bar{E}$, and then a second pass on the intermediate string B or $\bar{B}$ which must be stored temporarily. We do not need a separate pass to compute $\pi(B)$ or $\pi(\bar{B})$, since these values can be evaluated on-the-fly while computing B or $\bar{B}$. However, this really matters only for long inputs, where the benefits of redundancy-free encryption are less pronounced.

Any plaintext represented as a string of at least 2κ bits can be encrypted without requiring any special encoding, and without using any downstream symmetric-key cipher or other hybrid component. The ciphertext overhead is a single group element in $\mathbb{G}$.

3.4 Security Reduction

We prove the security of our scheme in the well-known and very standard sense of IND-CCA2 security, or indistinguishability under an adaptive chosen-ciphertext attack. The reduction will proceeds from an instance of the Gap-DH problem, in the random oracle model.

We recall that the Gap-DH problem is to solve the CDH problem given access to a DDH oracle. In a computational group $\mathbb{G}$, such an instance is a triple $(U, V, W) = (U, U^v, U^w) \in \mathbb{G}^3$, and the task is to compute the value $U^{vw} \in \mathbb{G}$, given repeated access to a decision oracle indicating whether an input tuple $(A, B, C, D) \in \mathbb{G}^4$ satisfies the relation $\mathrm{dlog}_A(B) = \mathrm{dlog}_C(D)$.

Theorem 1. *The miniature public-key cryptosystem is IND-CCA2 secure in the random oracle model, provided that the Gap Diffie-Hellman assumption holds in $\mathbb{G}$. The reduction is tight w.r.t. computational cost ("efficiency") and success probability ("efficacy") simultaneously.*

Proof. Suppose there is an adversary $\mathcal{A}$ that breaks the encryption scheme. We build from it an algorithm $\mathcal{B}$ that solves the Gap-DH problem by simulating an attack environment to such an adversary. During the course of the interaction, the simulator will record the answers it makes in response to all queries, and additionally maintain two separate "watch-lists" for Φ and Ψ.

Key generation. $\mathcal{B}$ is given access to a Decision Diffie-Hellman oracle $\mathcal{DDH} : \mathbb{G}^4 \to \{0, 1\}$; it receives a Diffie-Hellman instance $(U, V, W) = (U, U^v, U^w) \in \mathbb{G}^3$, and is to compute $U^{vw} \in \mathbb{G}$.

To start the simulation, $\mathcal{B}$ gives to $\mathcal{A}$ the public key $\mathsf{Pk} = (U, V)$, implicitly letting $\mathsf{Sk} = v$.

Decryption queries. $\mathcal{A}$ makes adaptive decryption queries on any ciphertexts $(D_k, E_k) \in \mathbb{G} \times \mathcal{M}$.

To respond, $\mathcal{B}$ sifts the query logs for a random oracle query $\Phi(D_j, C_j)$ such that $D_j = D_k$ and $C_j = {D_k}^{\mathsf{Sk}}$. To do this in constant time, $\mathcal{B}$ can maintain a hash-table of those oracle queries such that $\mathcal{DDH}(U, V, D_j, C_j) = 1$. Let thus (D_j, C_j) be the retrieved entry, if it exists.

- If it does, let $\phi_j = \Phi(D_k, C_j)$ be the previously assigned value; the simulator then computes $B_k \leftarrow E_k \oplus \phi_j$ and $A_k \leftarrow {C_j}^{\pi(B_k)}$, and returns $M_k \leftarrow B_k \oplus \Psi(A_k)$ as the plaintext.
- Otherwise, the simulator simply returns a random string $M_k \in_\$ \mathcal{M}$, while privately adding the triple (D_k, E_k, M_k) to the watch-list associated with Φ, for future use given below.

Hash-Φ queries. $\mathcal{A}$ adaptively queries the random oracle Φ on unique input pairs $(D_j, C_j) \in \mathbb{G}^2$.

To respond, $\mathcal{B}$ picks a random string $\phi_j \in_\$ \mathcal{M}$ which it returns as answer to the query. Additionally, it tests whether $\mathcal{DDH}(U, V, D_j, C_j) = 1$, in which case it pulls from the watch list associated with Φ all the triples (D_k, E_k, M_k) such that $D_k = D_j$. For all such triples, the simulator lets $B_k \leftarrow E_k \oplus \phi_j$, computes $A_k \leftarrow {C_j}^{\pi(B_k)}$, defines $\psi_k \leftarrow B_k \oplus M_k$, adds the pair (A_k, ψ_k) to the watch-list associated with Ψ, and deletes the triple from the list of Φ.

Observe that all E_k and thus all A_k are necessarily distinct, unless π collided, and that the work of the simulator is linear in the number of triples that were pulled from the watch-list. Later, we account for the small probability of getting a collision $A_{k_1} = A_{k_2}$ for $D_{k_1} \neq D_{k_2}$.

Hash-Ψ queries. $\mathcal{A}$ adaptively queries the random oracle Ψ on arbitrary unique inputs $A_i \in \mathbb{G}$.

To respond, $\mathcal{B}$ first determines whether the watch-list associated with Ψ contains a pair (A_k, ψ_k) with $A_k = A_i$. If there exists such a pair, the simulator removes it from the watch-list and returns the string ψ_k; otherwise, it returns a fresh random string $\psi_i \in_\$ \mathcal{M}$.

Challenge. $\mathcal{A}$ at some point outputs two messages M_1 and M_2 on which it wishes to be challenged.

To create the challenge, $\mathcal{B}$ picks a random string $E^* \in_\$ \mathcal{M}$, sets $D^* \leftarrow W$ from the Gap-DH instance, and declares the challenge ciphertext to be (D^*, E^*). It disregards M_1 and M_2.

Additional queries. $\mathcal{A}$ makes more adaptive decryption and random oracle queries on arbitrary inputs (but no decryption query on the challenge ciphertext), to which $\mathcal{B}$ responds as before.

As it services the queries, the simulator is now on the lookout for a query $\Phi(D^*, C^*)$ such that $D^* = W$ and $\mathcal{DDH}(U, V, W, C^*) = 1$. As soon as $\mathcal{A}$ makes this query, $\mathcal{B}$ terminates the simulation and outputs $C^* = U^{vw}$ as solution to the Gap-DH instance.

Outcome. If the adversary never asks for the value of $\Phi(W, U^{vw})$, its advantage must be zero, since then the simulation is perfect and the ciphertext is random. On the contrary, as soon as $\mathcal{A}$ makes this particular query, $\mathcal{B}$ obtains the solution it seeks without further interaction.

We now analyze the parameters of the reduction. We consider both *efficacy* (*i.e.*, the probability of success) and *efficiency* (*i.e.*, the computational overhead needed for a successful reduction).

Reduction Efficacy. It is easy to see that $\mathcal{B}$'s probability of solving Gap-DH is no less than $\mathcal{A}$'s advantage in the IND-CCA2 attack, minus a negligible loss $\Delta\epsilon$ that corresponds to the probability that the simulator made two conflicting random oracle assignments. A conflict can arise for $\Psi(A_k)$ due to a collision $A_{k_1} = {C_{j_1}}^{\pi(E_{k_1} \oplus \phi_{j_1})} = {C_{j_2}}^{\pi(E_{k_2} \oplus \phi_{j_2})} = A_{k_2}$ when $C_{j_1} \neq C_{j_2}$. Since the ϕ_j are jointly independent of the C_j and E_k, and since every troublesome C_j can be traced to a watch-list entry that in turn originates from a unique decryption query, the probability of such a collision over q_d decryption queries, which dictates the total efficacy loss of the system, is given by the birthday bound:

$$\Delta\epsilon = \epsilon(\mathcal{A}) - \epsilon(\mathcal{B}) \leq (q_d)^2/p \approx (q_d)^2\, 2^{-2\kappa} = \mathrm{negl}(\kappa) \ .$$

Reduction Efficiency. To express $\mathcal{B}$'s running time of in terms of $\mathcal{A}$'s, let us assume that the adversary makes q_d decryption and q_Φ and q_Ψ hash queries, and that each exponentiation in $\mathbb{G}$ or $\mathcal{DDH}$ query costs the simulator one time unit. The simulation time overhead $\Delta\tau$ is then given by $\Delta\tau = \tau(\mathcal{B}) - \tau(\mathcal{A}) = \Theta(q_d + q_\Phi + q_\Psi)$, from which we deduce that the running times of $\mathcal{A}$ and $\mathcal{B}$ are within a constant factor $\gtrsim 1$ (1 being the best possible ratio):

$$\tau(\mathcal{B})/\tau(\mathcal{A}) = \Theta(1) \ .$$

It follows that the reduction is tight in all parameters, as long as the number of random oracle and decryption queries made by the adversary is sub-exponential in κ, as required. $\square$

3.5 Practical Extensions

We briefly describe two simple extensions to the basic scheme, which we expect to be useful in certain applications.

3.5.1 Adaptive Chosen-Ciphertext Security *vs.* Integrity

Most CCA2-secure cryptosystems, with or without random oracles, achieve security against active attacks by performing an integrity check during the decryption process, based on some amount of redundancy that is embedded in the ciphertext during encryption. Cryptosystems of this kind include

Dolev-Dwork-Naor [17], Cramer-Shoup [14, 15], Fujisaki-Okamoto [19, 20], Kurosawa-Desmedt [30], and Canetti-Halevi-Katz [12], among others. Usually the redundancy is secret, but need not be.

By contrast, our scheme does *not* authenticate the ciphertext; it is similar in that respect to a few other systems such as Phan-Pointcheval [36] and Kurosawa-Matsuo [31] as already discussed. Indeed, without redundancy there cannot be a test to reject malformed ciphertexts, and thus the decryption process always succeeds. Hence there is no such thing as an "incorrect" ciphertext. (We remark, however, that because the IND-CCA2 security property implies PA-CCA2, or plaintext awareness, any ciphertext that was not created using the proper procedure will safely decrypt to an unpredictable and useless plaintext.)

In some applications, it may be desirable to detect that a ciphertext has been tampered with. One solution is of course to use a "traditional" efficient CCA2-secure scheme, such as Fujisaki-Okamoto in the random oracle model or Kurosawa-Desmedt in the standard model. Another solution is to add a small amount of redundancy in the plaintext of our scheme, such as a few zeros. This approach might be more desirable in cases where a quick and inexpensive integrity test is desired but not required for the security of the larger system: in this case adding a few zeros to the plaintext of our scheme will be the cheapest and most effective solution.

3.5.2 Non-interactive Distributed Threshold Decryption

Recall that in a threshold public-key system, a number of distributed "partial decryption centers" compute partial decryptions from the ciphertext, or shares, which are then combined in a threshold manner by a single combiner to produce the final plaintext; *cf.*, *e.g.*, [22].

As mentioned earlier, CCA2-secure threshold cryptosystems are difficult to deploy based on the two-key paradigm, and also using the random-oracle-based Fujisaki-Okamoto transformation, because the decryption process will require the partial decryptors to communicate with each other in order to decide whether a ciphertext is valid or not. Essentially, this is because the redundancy in those schemes is secret [38], which makes it difficult to perform a validity test before the plaintext has been recovered. By contrast, the identity-based approach is much more conducive to secure threshold decryption under active attacks, because its redundancy is public and can be checked non-interactively by the decryption centers without costly inter-communications [6].

Our scheme appears to be easy to turn into a non-interactive CCA2-secure threshold system. The reasons for this are twofold. First, since the security of our scheme does not depend on any integrity check, the difficulty of conducting such a check in a threshold setting should have no ill effect. Second, the algebra of the scheme itself turns out to be very propitious to secret sharing, because the secret key Sk is only used once in the decryption process, to compute $\bar{C} \leftarrow \bar{D}^{\mathsf{Sk}}$. Thus, our scheme can be used as a basis for a threshold scheme, by splitting the secret key Sk into a number of random shares $\mathsf{Sk}_1, ..., \mathsf{Sk}_n$ using Shamir's secret sharing. The decryption centers would use those shares to produce decryption shares (analogous to) $\bar{C}_i \leftarrow \bar{D}^{\mathsf{Sk}_i}$. With enough of those, the combiner could perform Lagrange interpolation "in the exponent" to recover the value of $\bar{C} = \prod_i \bar{C}_i^{\Lambda_i}$, where the Λ_i are publically computable Lagrange coefficients. Once it knew $\bar{C}$, the combiner could finish the decryption without further interacting with the decryption centers.

Of course, the above is only an outline of the central idea, since in addition we must prevent partial decryptions $\bar{C}_i$ from leaking damaging information; recall that in the non-threshold scheme the intermediate value $\bar{C}$ is never revealed. This is easy to do by adding safety checks to the decryption shares $\bar{C}_i$ by playing with D without interfering with the interpolation of $\bar{C} = \prod_i \bar{C}_i^{\Lambda_i}$. Unfortunately, this will increase the ciphertext overhead by at least one additional group element.

In Appendix A.1, we show as an example how to turn our scheme into a non-interactive threshold scheme based on the above ideas. It requires one element's worth of redundancy for CCA2-secure threshold decryption, but it can also be operated in non-threshold mode without the redundancy.

We call "strippable" a threshold scheme that has this removable redundancy feature, and is thus operable in one of two modes: threshold with redundancy, and direct without redundancy. The dual one-time pad structure of the underlying encryption will confer CCA2 after removal of the threshold redundancy as in Section 3.2 (albeit under a reduction to a different gap assumption, because a bilinear pairing is needed for the threshold mode; see Appendix A.1.)

It remains open to perform non-interactive threshold decryption securely against active attacks without using any redundancy (with only one group element of total ciphertext overhead).

3.6 Implementation on Curves

Although our scheme generally relies on the Gap-DH assumption, it is possible to implement it in a computational group $\mathbb{G}$ where DDH is known to be easy (and CDH still believed to be hard): in this case the $\mathcal{DDH}$ oracle can actually be implemented and locally computed from public information without making any oracle call, causing Gap-DH to reduce to the usual CDH assumption. In such groups, the security of the scheme thus follows from computational Diffie-Hellman, which has of course been studied extensively.

Elliptic curves equipped with an efficiently computable bilinear pairing are an obvious choice for the group $\mathbb{G}$, because the pairing lets us decide (but not compute) the Diffie-Hellman problem efficiently. (More precisely, $\mathbb{G}$ will be a prime-order subgroup of the group of points on a "pairing-friendly" curve.) We refer to [7] and the abundant literature on pairings for details; see also [9] for an informal tour, and [21] for a compendium of their cryptographic properties.

It follows that, on pairing-friendly curves, and more generally in any computational group with a bilinear map, the mere fact that the $\mathcal{DDH}$ oracle *could* be implemented efficiently, means that our scheme has a tight IND-CCA2 security reduction to the CDH assumption rather than Gap-DH, in the random-oracle model.

We emphasize that for this simplification to occur, the pairing only needs to exist; in reality we do not need to use it or implement it, or even know how to implement it. Consequently, an existential proof that there exists an efficient pairing (or of any other way to make the required DDH determination in the simulation) is sufficient to relax Gap-DH into the weaker CDH assumption.

4 Summary

We have proposed a very simple public-key cryptosystem with the most compact ciphertext for a given level of exact CCA2 security, without relying on hybrid constructions. Earlier constructions with similarly compact ciphertexts required complex modes of operations for block ciphers and/or stronger assumptions.

The ciphertext appears difficult to shrink further because it has no redundancy left to remove, and also because the scheme offers a tight security reduction (both efficacy-wise and efficiency-wise) to a basic complexity assumption (Gap-DH, or just CDH if the arithmetic is done on a pairing-friendly curve). Although our ciphertext still contains an apparent excess of randomness, like all Discrete-Log-based encryption schemes, it appears difficult to remove any of it without departing from the algebraic realm of prime-order computational groups and the hardness of Discrete Log.

We have utilized a few new tricks to achieve "direct" tightness without redundancy. These tricks are set in the random oracle model, but we managed to avoid one of the main problems associated with the random oracle methodology, namely, the fact that, once instantiated, the hash function can be queried offline a practically unlimited number of times by any sufficiently powerful attacker. Since our scheme's security is not sensitive to the number of queries (beneath the birthday bound), exact security remains tight as long as the hash function is adequately modeled as a black box.

Of course, it would be nice to construct a redundancy-free CCA2-secure public-key encryption system in the standard model (even with a polynomially sloppy security reduction). However, this appears to be a very difficult problem, because, without redundancy in the ciphertext, it is not clear how the simulator could feasibly extract the information needed to answer decryption queries. In this respect, our scheme represents another in a long series of *a priori* surprising results that crucially rely on the random oracle methodology [2, 13, 26, 18].

We hope that our scheme will appeal to the practitioners of cryptography. Ideal uses for it would be in radio systems that make frequent transmissions of short independent messages, or, more generally, in all bandwidth-contrained environments where active attacks are a concern.

Acknowledgements

The author thanks the anonymous referees of Asiacrypt 2007 for useful comments, and Eike Kiltz for helpful pointers and a discussion that shaped the motivational exposition of Section 3.1.

References

[1] Michel Abdalla, Mihir Bellare, and Phil Rogaway. The oracle Diffie-Hellman assumption and an analysis of DHIES. In *Topics in Cryptology—CT-RSA 2001*, volume 2020 of *LNCS*, pages 143–58. Springer, 2001.

[2] Mihir Bellare and Phil Rogaway. Random oracle are practical: A paradigm for designing efficient protocols. In *ACM Conference on Computer and Communications Security—CCS 2003*, pages 62–73. ACM Press, 1993.

[3] Mihir Bellare and Phil Rogaway. Optimal asymmetric encryption - how to encrypt with RSA. In *Advances in Cryptology—EUROCRYPT 1994*, volume 950 of *LNCS*, pages 92–111. Springer, 1995.

[4] Dan Boneh. Simplified OAEP for the RSA and Rabin functions. In *Advances in Cryptology—CRYPTO 2001*, volume 2139 of *LNCS*, pages 275–91. Springer, 2001.

[5] Dan Boneh and Xavier Boyen. Efficient selective-ID secure identity based encryption without random oracles. In *Advances in Cryptology—EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 223–238. Springer, 2004.

[6] Dan Boneh, Xavier Boyen, and Shai Halevi. Chosen ciphertext secure public key threshold encryption without random oracles. In *Topics in Cryptology—CT-RSA 2006*, volume 3860 of *Lecture Notes in Computer Science*, pages 226–43. Springer, 2006.

[7] Dan Boneh and Matt Franklin. Identity-based encryption from the Weil pairing. In *Advances in Cryptology—CRYPTO 2001*, volume 2139 of *LNCS*, pages 213–29. Springer, 2001.

[8] Dan Boneh and Jonathan Katz. Improved efficiency for CCA-secure cryptosystems built using identity based encryption. In *Topics in Cryptology—RSA-CT 2005*, LNCS. Springer, 2005.

[9] Xavier Boyen. A promenade through the new cryptography of bilinear pairings. In *IEEE Information Theory Workshop—ITW 2006*, pages 19–23. IEEE Press, 2006. Available at `http://www.cs.stanford.edu/~xb/itw06/`.

[10] Xavier Boyen, Qixiang Mei, and Brent Waters. Direct chosen ciphertext security from identity-based techniques. In *ACM Conference on Computer and Communications Security—CCS 2005*, pages 320–9. ACM Press, 2005.

[11] Ran Canetti and Shafi Goldwasser. An efficient threshold public key cryptosystem secure against adaptive chosen ciphertext attack. In *Advances in Cryptology—EUROCRYPT 1999*, LNCS, pages 90–106. Springer, 1999.

[12] Ran Canetti, Shai Halevi, and Jonathan Katz. Chosen-ciphertext security from identity-based encryption. In *Advances in Cryptology—EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 207–22. Springer, 2004.

[13] Jean-Sébastien Coron. On the exact security of full-domain-hash. In *Advances in Cryptology—CRYPTO 2000*, volume 1880 of *LNCS*, pages 229–35. Springer, 2000.

[14] Ronald Cramer and Victor Shoup. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In *Advances in Cryptology—CRYPTO 1998*, volume 1462 of *LNCS*. Springer, 1998.

[15] Ronald Cramer and Victor Shoup. Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption. In *Advances in Cryptology—EUROCRYPT 2002*, volume 2729 of *LNCS*, pages 45–64. Springer, 2002.

[16] Ronald Cramer and Victor Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal of Computing*, 33:167–226, 2003.

[17] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-malleable cryptography. In *ACM Symposium on Theory of Computing—STOC 1991*, pages 542–52. ACM Press, 1991.

[18] Marc Fischlin. Communication-efficient non-interactive proofs of knowledge with online extractors. In *Advances in Cryptology—CRYPTO 2005*, LNCS, pages 152–68. Springer, 2005.

[19] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In *Advances in Cryptology—CRYPTO 1999*, LNCS, pages 537–54. Springer, 1999.

[20] Eiichiro Fujisaki and Tatsuaki Okamoto. How to enhance the security of public-key encryption at minimum cost. *IEICE Transactions on Fundamentals*, E83-9(1):24–32, 2000.

[21] Steven Galbraith, Kenneth Paterson, and Nigel Smart. Pairings for cryptographers. Cryptology ePrint Archive, Report 2006/165, 2006. `http://eprint.iacr.org/`.

[22] Rosario Gennaro, Tal Rabin, Stanisław Jarecki, and Hugo Krawczyk. Robust and efficient sharing of RSA functions. *Journal of Cryptology*, 13(2):273–300, 2000.

[23] Shai Halevi and Phil Rogaway. A tweakable enciphering mode. In *Advances in Cryptology—CRYPTO 2003*, volume 2729 of *LNCS*, pages 482–99. Springer, 2003.

[24] Shai Halevi and Phil Rogaway. A parallelizable enciphering mode. In *Topics in Cryptology—CT-RSA 2004*, LNCS, pages 292–304. Springer, 2004.

[25] Antoine Joux and Kim Nguyen. Separating decision Diffie-Hellman from computational Diffie-Hellman in cryptographic groups. *Journal of Cryptology*, 16(4):239–47, 2003.

[26] Jonathan Katz and Nan Wang. Efficiency improvements for signature schemes with tight security reductions. In *ACM Conference on Computer and Communications Security—CCS 2003*, pages 155–64. ACM Press, 2003.

[27] Eike Kiltz. Chosen-ciphertext security from tag-based encryption. In *Proceedings of TCC 2006*, LNCS, pages 581–600. Springer, 2006.

[28] Eike Kiltz. Chosen-ciphertext secure key encapsulation based on hashed gap decisional Diffie-Hellman. In *Public Key Cryptography—PKC 2007*, volume 4450 of *LNCS*, pages 282–97. Springer, 2007.

[29] Eike Kiltz and Gregory Neven. Hedging random oracles with generic groups. Unpublished, 2007.

[30] Kaoru Kurosawa and Yvo Desmedt. A new paradigm of hybrid encryption scheme. In *Advances in Cryptology—CRYPTO 2004*, LNCS, pages 426–42. Springer, 2004.

[31] Kaoru Kurosawa and Toshihiko Matsuo. How to remove MAC from DHIES. In *Proceedings of ACISP 2004*, volume 3108 of *LNCS*, pages 236–47. Springer, 2004.

[32] Benoît Libert and Jean-Jacques Quisquater. Identity based encryption without redundancy. In *Proceedings of ACNS 2005*, volume 3531 of *LNCS*, pages 285–300. Springer, 2005.

[33] Xianhui Lu, Xuejia Lai, and Dake He. Efficient chosen ciphertext secure PKE scheme with short ciphertext. Cryptology ePrint Archive, Report 2007/210, 2007. `http://eprint.iacr.org/`.

[34] Alfred J. Menezes, Paul C. Van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography.* CRC Press, 1997.

[35] Moni Naor and Moti Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *ACM Symposium on Theory of Computing—STOC 1990*, pages 427–37. ACM Press, 1990.

[36] Duong Hieu Phan and David Pointcheval. Chosen-ciphertext security without redundancy. In *Advances in Cryptology—ASIACRYPT 2003*, volume 2894 of *LNCS*, pages 1–18. Springer, 2003.

[37] Duong Hieu Phan and David Pointcheval. OAEP 3-round: A generic and secure asymmetric encryption padding. In *Advances in Cryptology—ASIACRYPT 2004*, volume 3329 of *LNCS*. Springer, 2004.

[38] Victor Shoup and Rosario Gennaro. Securing threshold cryptosystems against chosen ciphertext attack. *Journal of Cryptology*, 15(2):75–96, 2002.

A Appendix

A.1 A "Strippable" Non-Interactive Threshold Scheme

In this section we build a non-interactive CCA2-secure threshold encryption scheme based on the strategy outlined in Section 3.5.2. First, it is instructive to see why extra safeguards are needed.

Suppose we merely applied Lagrange interpolation to the scheme of Section 3, and used for decryption shares the intermediate values $\bar{C}_i \leftarrow D^{\mathsf{sk}_i}$. An active adversary wishing to decrypt some ciphertext (D, E) could make a decryption query on $(D^s, E') \neq (D, E)$ for some blinding exponent $s \in_\$ \mathbb{F}_p$; it would obtain the shares $C_i^s \leftarrow (D^s)^{\mathsf{sk}_i}$ in return, and be able to reconstruct $C^s = (D^{\mathsf{sk}})^s$ and thus $C = D^{\mathsf{sk}}$. From there, the original ciphertext would be easy to decrypt, illegitimately. Protecting the decryption shares $\bar{C}_i$ from mauling is thus a necessity. It is also relatively easy to do, especially in the random-oracle model, without changing the structure of our underlying scheme. One idea is to add a (thin) identity-based encryption layer on top of the existing structure, based on preexisting randomness. Its purpose will be to make the ciphertext "tamper-evident", even to those who are not aware of the private key, such as the partial decryption servers.

For this, we can use the BB1 scheme [5] and recycle the randomness already present in the "miniature" ciphertext. This setup has the advantage of preserving the tightness of the security reduction, while only adding one group element's worth of redundancy to the ciphertext. Of course, the main drawback of using a pairing-based IBE subsystem is that the pairing is no longer optional. The following is an example of such a "miniature-&-threshold" hybrid construction:

Context: Same as in the main scheme, except that $\mathbb{G}$ is now equipped with an efficient bilinear pairing $e : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_t$. We use the following hash functions:

- Two collision-resistant hash functions $\pi : \mathcal{M} \to \mathbb{F}_p^\times$ and $\theta : \mathbb{G} \times \mathcal{M} \to \mathbb{F}_p$.
- Two "random-oracle" hash functions $\Phi : \mathbb{G}_t \times \mathbb{G} \to \mathcal{M}$ and $\Psi : \mathbb{G}_t \to \mathcal{M}$.

Key generation: Draw a random exponent $s \in_\$ \mathbb{F}_p^\times$, and calculate $Y = e(U,U)^s$. Also draw an additional random generator $U' \in_\$ \mathbb{G}$, obtained, *e.g.*, by raising U to a random power.

The public encryption key is $\mathsf{Pk} \leftarrow (U,\ U',\ Y) \in \mathbb{G}^2 \times \mathbb{G}_t$.
The private decryption key is $\mathsf{Sk} \leftarrow s \in \mathbb{F}_p^\times$.

Each center receives a private key share $\mathsf{Sk}_i \leftarrow s_i$, where the s_i are random shares of s in Shamir's polynomial secret sharing scheme with publically computable Lagrange coefficients. Additionally, the verification values $Y_i \leftarrow e(U,U)^{s_i}$ are made public.

Encryption: Given Pk and a plaintext $\mathsf{Msg} \in \mathcal{M}$, pick a randomizer $r \in_\$ \mathbb{F}_p^\times$, and let,

$$\begin{aligned}
&A \leftarrow Y^r \\
&B \leftarrow \Psi(A) \oplus \mathsf{Msg} \\
&C \leftarrow Y^{r/\pi(B)} \\
&D \leftarrow U^{r/\pi(B)} \\
&E \leftarrow \Phi(C,D) \oplus B \\
&F \leftarrow (U^{\theta(D,E)}V)^{-r/\pi(B)}
\end{aligned}$$

The ciphertext is $\mathsf{Ctx} = (D,\ E,\ F) \in \mathbb{G} \times \mathcal{M} \times \mathbb{G}$.

Partial Decryption (@ i-th center): Given a ciphertext $\mathsf{Ctx} = (\bar{D},\ \bar{E},\ \bar{F})$, and a private key share $\mathsf{Sk}_i = s_i$, verify that $e(\bar{D},\ U^{\theta(\bar{D},\bar{E})}V)e(\bar{F},\ U) = 1$, and, if so, pick $t \in_\$ \mathbb{F}_p$, and let,

$$\begin{aligned}
&K_i \leftarrow (U^{\theta(\bar{D},\bar{E})}V)^t W_i \\
&L_i \leftarrow U^{-t}
\end{aligned}$$

The partial decryption share is $\mathsf{Ctx}_i = (K_i, L_i) \in \mathbb{G}^2$.

Final Decryption (@ combiner): Given a ciphertext $\mathsf{Ctx} = (\bar{D},\ \bar{E},\ \bar{F})$, and enough decryption shares $\mathsf{Ctx}_i = (\bar{K}_i, \bar{L}_i)$ from distinct centers, check that $e(\bar{D}, U^{\theta(\bar{D},\bar{E})}V)e(\bar{F}, U) = 1$, and that $e(\bar{K}_i, U)e(\bar{L}_i, U^{\theta(\bar{D},\bar{E})}V) = Y_i$ for each share i. If so, determine the interpolation coefficients Λ_i for the selected set of shares, and let,

$$\begin{aligned}
&K \leftarrow \textstyle\prod_i (\bar{K}_i)^{\Lambda_i} \\
&L \leftarrow \textstyle\prod_i (\bar{L}_i)^{-\Lambda_i} \\
&\bar{C} \leftarrow e(\bar{D}, K)e(\bar{F}, L) \qquad \text{or equivalently} \qquad \bar{C} \leftarrow \textstyle\prod_i \bar{C}_i^{\Lambda_i} \quad \text{where} \quad \bar{C}_i \leftarrow e(\bar{D}, \bar{K}_i)/e(\bar{F}, \bar{L}_i) \\
&\bar{B} \leftarrow \bar{E} \oplus \Phi(\bar{C}, \bar{D}) \\
&\bar{A} \leftarrow \bar{C}^{\pi(\bar{B})} \\
&\bar{M} \leftarrow \bar{B} \oplus \Psi(\bar{A})
\end{aligned}$$

The decrypted plaintext is $\bar{\mathsf{Msg}} = \bar{M} \in \mathcal{M}$.

Decryption (directly from ciphertext, without threshold): Given the main private key Sk and a ciphertext $\mathsf{Ctx} = (\bar{D}, \bar{E}, \star)$, where only the first two ciphertexts components matter, verify that $1 \neq \bar{D} \in \mathbb{G}$, and, if so, let,

$$\begin{aligned}
&\bar{C} \leftarrow e(\bar{D}, U^{\mathsf{Sk}}) \qquad \text{comparable to an ``exponentiation into } \mathbb{G}_t\text{'' given the constant } U^{\mathsf{Sk}} \\
&\bar{B} \leftarrow \bar{E} \oplus \Phi(\bar{C}, \bar{D}) \\
&\bar{A} \leftarrow \bar{C}^{\pi(\bar{B})} \\
&\bar{M} \leftarrow \bar{B} \oplus \Psi(\bar{A})
\end{aligned}$$

The decrypted plaintext is $\bar{\mathsf{Msg}} = \bar{M} \in \mathcal{M}$.

We note that a threshold encryption system based on a similar "self-authenticating" IBE layer has been constructed before [6], in the standard model. Both systems require a pairing, and incur the same ciphertext overhead (two elements of $\mathbb{G}$ on top of the blinded message itself). They both have a tight security reduction to the computational Bilinear Diffie-Hellman (BDH) assumption defined in [7, 25]. The main difference with the scheme of [6] and other efficient threshold systems is that, here, the redundancy is only needed during threshold decryption, and can otherwise be removed *ex post facto* by truncating the ciphertext (D, E, F) into (D, E), in a process that we call "stripping".

Security-wise, our "strippable" threshold scheme satisfies two properties concurrently:

- With the full ciphertext (D, E, F), the scheme has a tight reduction to the BDH assumption, in the random-oracle model, for threshold and non-threshold CCA2 adversaries alike. Apart from the presence of random oracles, the reduction is essentially the same as in [6].
- With the stripped ciphertext (D, E), there is a tight reduction to the Gap-BDH assumption, in the random-oracle model, for non-threshold CCA2 adversaries only. This reduction exploits the "dual one-time pad" structure of the sequential blinding $\mathsf{Msg} \mapsto B \mapsto E$, as in Section 3.4.

"Strippable" threshold systems are useful in applications where the primary requirement is that ciphertexts be as compact as possible, and the secondary consideration that threshold decryption remain available for occasional use, at little or no extra cost to non-threshold recipients.