Public Key Trace and Revoke Scheme Secure against Adaptive Chosen Ciphertext Attack*

Yevgeniy Dodis Nelly Fazio

Computer Science Department
Courant Institute of Mathematical Science
New York University, USA
{dodis,fazio}@cs.nyu.edu

Abstract

A (public key) Trace and Revoke Scheme combines the functionality of broadcast encryption with the capability of traitor tracing. Specifically, (1) a trusted center publishes a single public key and distributes individual secret keys to the users of the system; (2) anybody can encrypt a message so that all but a specified subset of "revoked" users can decrypt the resulting ciphertext; and (3) if a (small) group of users combine their secret keys to produce a "pirate decoder", the center can trace at least one of the "traitors" given access to this decoder.

We construct the first *chosen ciphertext* (CCA2) secure Trace and Revoke Scheme based on the DDH assumption. Our scheme is also the first *adaptively secure* scheme, allowing the adversary to corrupt players at any point during execution, while prior works (e.g., [19, 21]) only achieves a very weak form of non-adaptive security even against chosen plaintext attacks. In fact, no CCA2 scheme was known even in the symmetric setting.

Of independent interest, we present a slightly simpler construction that shows a "natural separation" between the classical notion of CCA2 security and the recently proposed [20, 1] relaxed notion of gCCA2 security.

1 Introduction

A *broadcast encryption* scheme allows the sender to securely distribute data to a dynamically changing set of users over an insecure channel. Namely, it should be possible to selectively exclude (i.e., "revoke") a certain subset of users from receiving the data. For that reason, it is often convenient to think of broadcast encryption as a *revocation scheme*, since the revocation ability is what makes the task of broadcast encryption non-trivial. In particular, each user should receive an individualized decoder (i.e., a decryption device with a unique secret key) which decrypts only the ciphertexts intended for the given user. Broadcast encryption has numerous applications, including pay-TV systems, distribution of copyrighted material, streaming audio/video and many others.

The formal study of broadcast encryption was initiated by Fiat and Naor [11], who showed a scheme with message overhead roughly $O(z^2 \log^2 z \log N)$, where z is the maximum number of excluded users (so called *revocation threshold*) and N is the total number of users. Subsequent works include [16, 15, 13], and, more recently, [18, 14] which show how to achieve linear message overhead $O(z)$ and $\omega(\log N)$ storage per user.

*Extended version of [10].

A related line of work concerns *multicast security* [22, 17, 23, 4, 5]. However, in this setting revoking a single user involves changing the keys for all the users, which makes it inapplicable to situations where the receivers are "stateless", do not always stay "on-line", or where the set of receivers can change rapidly.

Most of the above works primarily concentrate on the centralized setting, where only the trusted center (the entity who generates all the secret keys) can send messages to the receivers. In the *public key* setting, studied in this paper, the center also prepares a fixed public key which allows any entity to play the role of the sender. Aside from achieving this extra functionality, the public key setting also allows the center to store secret keys in a more secure place than the station used for data transmission (e.g., off-line), and access this storage only for "system maintenance" (e.g., when a new user joins the system).

In the public key setting, the only known Broadcast Encryption Schemes have been constructed by [19, 21] based on the DDH assumption, and achieve public key and message overhead $O(z)$. In fact, these schemes are essentially identical: in the following we will refer to the work of [21], who emphasize more the public key nature of their scheme.

Concurrently with the present work, Dodis and Fazio [9] extended the efficient scheme of [18] to the asymmetric setting. The resulting public key Broadcast Encryption Scheme achieves constant key size, while maintaining similar ciphertext expansion, but does not enjoy full CCA2 security: in fact, it seems hard to obtain such a high level of security within the Subset Cover framework of [18].

Some Criticism. Despite providing a simple and elegant scheme, the work of [21] has several noticeable shortcomings. First, the given (informal) notion of security does not address the peculiar features of the revocation setting. Indeed, to show the "security" of revocation, [21] shows the following two claims: (1) the scheme is semantically secure when no users are revoked; (2) no set of z a-priori fixed users can compute the secret key of another user. Clearly, these properties do not imply the security notion we really care about and which informally states: (3) if the adversary controls some set $\mathcal{R}$ of up to z *revoked* users, then the scheme remains semantically secure. Actually, the scheme of [21] can be shown to satisfy the needed property (3) only when the set $\mathcal{R}$ is chosen by the adversary *non-adaptively*, and in fact only if it is chosen before the adversary learns the public key. Such weak non-adaptive security is clearly insufficient for realistic usages of a public key revocation scheme.

Most importantly, the extended scheme of [21] is proven to be CCA2-secure when none of the users is corrupted, but stops being such the moment just a single user is corrupted, even if this user is immediately revoked for the rest of the protocol. Again, this is too weak — the scheme should remain CCA2-secure *even after many users have been revoked*. As we will see, achieving this strong type of security is very non-trivial, and requires a much more involved scheme than the one proposed by [21].

Our Contributions. We introduce for the first time a precise formalization of an appropriate notion of adaptive security for public key Broadcast Encryption Schemes, for both the CPA and the CCA2 setting, which naturally models property (3) mentioned above. We construct the first *adaptive chosen ciphertext* (CCA2) secure public key Broadcast Encryption Scheme under the DDH assumption (with no random oracles). We remark that no CCA2-secure schemes were known even in the symmetric setting. Moreover, it doesn't seem obvious how to extend current symmetric schemes (e.g. [18]) to meet the CCA2 notion. Our public key scheme is based on the regular Cramer-Shoup encryption [7, 8], but our extension is non-trivial, as we have to resolve some difficulties inherent to the Broadcast Encryption setting. Our CCA2-secure scheme requires a constant user storage and a

public key size proportional to the revocation threshold z. The length of each ciphertext, and the time to encrypt and decrypt a message are all proportional to $O(z)$.

As a preliminary step, we show how to modify the CPA-scheme of [21] to achieve a much more appropriate notion of adaptive security, while maintaining essentially the same efficiency in all the parameters (up to a factor of 2).

Of independent interest, we also provide another scheme achieving a slightly weaker (but still very strong) notion of *generalized* CCA2 security (gCCA2) [20, 1]. As argued in [1], the gCCA2 security is much more robust to syntactic changes, while still sufficient for all known uses of CCA2 security. Interestingly, all the examples separating CCA2- and gCCA2-secure encryption were "artificial" in a sense that they made a more complicated scheme from an already existing CCA2-secure encryption. Our work shows the first "natural" separation, but for the setting of *broadcast* public key encryption.

A Note on Traitor Tracing. As first explicitly noticed by Gafni et al. [12], Broadcast Encryption is most useful when combined with a *Traitor Tracing* mechanism [6] by which the center can extract the identity of (at least one) "pirate" from any illegal decoder produced combining decryption equipments of a group of legal members (the "traitors"). By slightly modifying standard tracing algorithms from previous weaker schemes (e.g. [19, 21]), tracing algorithms can be added to our schemes, thus yielding fully functional *Trace and Revoke* schemes [19]. However, we will focus only on Broadcast Encryption (i.e. revocation), which is also the main novelty of this paper.

2 Notations and Basic Facts

Lagrange Interpolation in the Exponent. Let q be a prime and $f(x)$ a polynomial of degree z over $\mathbb{Z}_q$; let $j_0, \ldots, j_z$ be distinct elements of $\mathbb{Z}_q$, and let $f_0 = f(j_0), \ldots, f_z = f(j_z)$. Using Lagrange Interpolation, we can express the polynomial as $f(x) = \sum_{t=0}^{z}(f_t \cdot \lambda_t(x))$, where $\lambda_t(x) = \prod_{0 \le i \ne t \le z} \frac{j_i - x}{j_i - j_t}, t = 0, \ldots, z$. We can now define the Lagrange Interpolation Operator as follows:

$$\mathsf{LI}(j_0, \ldots, j_z; f_0, \ldots, f_z)(x) \doteq \sum_{t=0}^{z}(f_t \cdot \lambda_t(x)).$$

Now, consider any cyclic group $\mathbb{G}$ of order q and a generator g of $\mathbb{G}$. For any distinct values $j_0, \ldots, j_z$ of $\mathbb{Z}_q$ and (non necessarily distinct) elements $v_0, \ldots, v_z$ of $\mathbb{G}$, let us define the Lagrange Interpolation Operator in the Exponent as:

$$\mathsf{EXP\text{-}LI}(j_0, \ldots, j_z; v_0, \ldots, v_z)(x) \doteq g^{\mathsf{LI}(j_0,\ldots,j_z;\log_g v_0,\ldots,\log_g v_z)(x)} = \prod_{t=0}^{z} g^{(\log_g v_t \cdot \lambda_t(x))} = \prod_{t=0}^{z} v_t^{\lambda_t(x)}.$$

The last expression shows that the function EXP-LI is poly-time computable, despite being defined in terms of discrete logarithms (which are usually hard to compute). We also remark on another useful property of the above operator: $\mathsf{EXP\text{-}LI}(j_0, \ldots, j_z; v_0^r, \ldots, v_z^r)(x) = [\mathsf{EXP\text{-}LI}(j_0, \ldots, j_z; v_0, \ldots, v_z)(x)]^r$. In what follows, we will refer to a function of the form $g^{f(x)}$, where $f(x)$ is a polynomial, as an EXP-polynomial.

DDH Assumption. The security of our schemes will rely on the Decisional Diffie-Hellman (DDH) Assumption in the group $\mathbb{G}$: namely, it is computationally hard to distinguish a random tuple (g_1, g_2, u_1, u_2) of four independent elements in $\mathbb{G}$ from a random tuple satisfying $\log_{g_1} u_1 = \log_{g_2} u_2$ (for a survey, see [3]).

A Probabilistic Lemma. The following useful lemma states that to estimate the difference between two related experiments U_1 and U_2, it is sufficient to bound the probability of some event F which "subsumes" all the differences between the experiments. Formally,

Lemma 1 *If U_1, U_2 and F are events such that $(U_1 \wedge \neg F)$ and $(U_2 \wedge \neg F)$ are equivalent events, then $\left|\Pr[U_1] - \Pr[U_2]\right| \leq \Pr[F]$.*

3 Definition of Broadcast Encryption Scheme

Since a public-key broadcast encryption is typically used by encrypting a session key s for the privileged users (this encryption is called the *enabling block*), and then symmetrically encrypting the "actual" message with s, we will often say that the goal of a Broadcast Encryption Scheme is to *encapsulate* [8] a session key s, rather than to encrypt a message M.

Definition 2 (Broadcast Encryption Scheme)
A Broadcast Encryption Scheme BE *is a 4-tuple of* poly*-time algorithms* (KeyGen, Reg, Enc, Dec)*, where:*

- KeyGen*, the* key generation algorithm*, is a probabilistic algorithm used by the center to set up all the parameters of the scheme.* KeyGen *takes as input a security parameter 1^λ and a revocation threshold z (i.e. the maximum number of users that can be revoked) and generates the public key PK and the master secret key SK_{BE}.*
- Reg*, the* registration algorithm*, is a probabilistic algorithm used by the center to compute the secret initialization data needed to construct a new decoder each time a new user subscribes to the system.* Reg *receives as input the master key SK_{BE} and a (new) index i associated with the user; it returns the user's secret key SK_i.*
- Enc*, the* encryption algorithm*, is a probabilistic algorithm used to encapsulate a given session key s within an enabling block $\mathcal{T}$.* Enc *takes as input the public key PK, the session key s and a set $\mathcal{R}$ of revoked users (with $|\mathcal{R}| \leq z$) and returns the enabling block $\mathcal{T}$.*
- Dec*, the* decryption algorithm*, is a deterministic algorithm that takes as input the secret key SK_i of user i and the enabling block $\mathcal{T}$ and returns the session key s that was encapsulated within $\mathcal{T}$ if i was a legitimate user when $\mathcal{T}$ was constructed, or the special symbol $\perp$.*

3.1 Security of Revocation

Intuitively, we would like to say that even if a malicious *adversary* $\mathcal{A}$ learns the secret keys of at most z users, and these users are later revoked, then subsequent broadcasts do not leak any information to such adversary. The security threat posed by such adversary is usually referred to as *Chosen Plaintext Attack* (CPA), and a Broadcast Encryption Scheme withstanding such an attack is said to be *z-Resilient against* CPA. It is well known that such an attack is not powerful enough to model some realistic adversarial scenarios, e.g. in the presence of an insider who helps the adversary to get decryptions of arbitrary ciphertexts.

To be on the safe side, it is possible to consider the *Chosen Ciphertext Attack* (CCA2) in which the adversary is allowed to "play" with the decryption machinery as she wishes, subject only to the condition that she doesn't ask about enabling blocks closely related to her "challenge" $\mathcal{T}^*$. In

formalizing the notion of "close relationship", the usual treatment is to impose a minimal restriction to the adversary, just disallowing her to submit the challenge itself to the decryption machinery. As already noted in [20, 1], such a mild constraint does in turn restrict too much the class of schemes that can be proven secure, excluding even schemes that ought to be considered secure under a more intuitive notion. For this reason, it seems more reasonable to consider a variant of the $\mathsf{CCA2}$, to which we will refer to as *Generalized Chosen Ciphertext Attack* ($\mathsf{gCCA2}$), following the terminology introduced in [1].

In a Generalized Chosen Ciphertext Attack, the set of enabling blocks the adversary is forbidden to ask about is defined in term of an efficiently computable equivalence relation $\Re(\cdot,\cdot)$. In fact, in the case of a broadcast (as opposed to ordinary) encryption, there is no unique decryption machinery, since the decryption algorithm can be used with the secret key of any legitimate user. For this reason, in our setting we need to consider a *family* of efficient equivalence relations $\{\Re_i(\cdot,\cdot)\}$, one for each user i. As in the regular case [1], the equivalence relation $\Re_i(\cdot,\cdot)$ corresponding to each user i needs to be *i-decryption-respecting*: equivalent enabling blocks under $\Re_i$ are guaranteed to have exactly the same decryption according to the secret data of user i. Finally, this family should form an *explicit* parameter of the scheme (i.e., one has to specify some decryption-respecting family $\{\Re_i\}$ when proving the $\mathsf{gCCA2}$ security of a given scheme).

Formal Model. We now formalize the above attack scenarios, starting with the CPA.

First, $(PK, SK_{\mathsf{BE}}) \leftarrow \mathsf{BE.KeyGen}(1^\lambda, z)$ is run and the adversary $\mathcal{A}$ is given the public key PK. Then $\mathcal{A}$ enters the user corruption stage, where she is given oracle access to the *User Corruption Oracle* $\mathsf{Cor}_{SK_{\mathsf{BE}}}(\cdot)$. This oracle receives as input the index i of the user to be corrupted, computes $SK_i \leftarrow \mathsf{BE.Reg}(SK_{\mathsf{BE}}, i)$ and returns the user's secret key SK_i. This oracle can be called *adaptively* for at most z times. Let us say that at the end of this stage the set $\mathcal{R}$ of at most z users is corrupted.

In the second stage, a random bit σ is chosen, and $\mathcal{A}$ can query the *Encryption Oracle* (sometimes also called the *left-or-right* oracle) $\mathcal{E}_{PK,\mathcal{R},\sigma}(\cdot,\cdot)$ on any pair of session keys s_0, s_1.[1] This oracle returns $\mathsf{Enc}(PK, s_\sigma, \mathcal{R})$. Without loss of generality (see [2]), we can assume that the encryption oracle is called exactly once, and returns to $\mathcal{A}$ the *challenge enabling block* $\mathcal{T}^*$. At the end of this second stage, $\mathcal{A}$ outputs a bit σ^* which she thinks is equal to σ. Define the *advantage* of $\mathcal{A}$ as $\mathsf{Adv}^{\mathsf{CPA}}_{\mathsf{BE},\mathcal{A}}(\lambda) \doteq |\Pr(\sigma^* = \sigma) - \frac{1}{2}|$.

Additionally, in the case of a Chosen Ciphertext Attack (generalized or not), $\mathcal{A}$ has also access to a *Decryption Oracle* $\mathcal{D}_{SK_{\mathsf{BE}}}(\cdot,\cdot)$, which she can query on any pair $\langle i, \mathcal{T}\rangle$, where i is the index of some user and $\mathcal{T}$ is any enabling block of her choice. $\mathcal{A}$ can call this oracle at any point during the execution (i.e., both in the first and in the second stage, arbitrarily interleaved with her other oracle calls). To prevent the adversary from directly decrypting her challenge $\mathcal{T}^*$, the decryption oracle first checks whether $\Re_i(\mathcal{T}, \mathcal{T}^*)$ holds[2]: if so, $\mathcal{D}$ outputs $\perp$; if not, $\mathcal{D}$ computes $SK_i \leftarrow \mathsf{BE.Reg}(SK_{\mathsf{BE}}, i)$ and uses it to output $\mathsf{BE.Dec}(i, \mathcal{T})$. As before, we define the corresponding advantages $\mathsf{Adv}^{\mathsf{gCCA2}}_{\mathsf{BE},\mathcal{A}}(\lambda)$ and $\mathsf{Adv}^{\mathsf{CCA2}}_{\mathsf{BE},\mathcal{A}}(\lambda)$.

Definition 3 (z-Resilience of a Broadcast Encryption Scheme)
Let $\mu \in \{\mathsf{CPA}, \mathsf{gCCA2}, \mathsf{CCA2}\}$. We say that a Broadcast Encryption Scheme BE is z-resilient against a μ-type attack if the advantage, $\mathsf{Adv}^{\mu}_{\mathsf{BE},\mathcal{A}}(\lambda)$, of any probabilistic poly-time algorithm $\mathcal{A}$ is a negligible function of λ.

[1] For the sake of generality, we could have allowed $\mathcal{A}$ to interleave the calls to $\mathsf{Cor}_{SK_{\mathsf{BE}}}(i)$ and $\mathcal{E}_{PK,\mathcal{R},\sigma}$ (where $\mathcal{A}$ can choose any i's and $\mathcal{R}$'s only subject to $i \notin \mathcal{R}$). However, this clumsier definition is easily seen to be equivalent to the one we present.

[2] This preliminary check applies to the standard Chosen Ciphertext Attack as well, which corresponds to all the $\Re_i$'s being the equality relation.

4 Revocation Schemes

In this section, we present three Broadcast Encryption Schemes, achieving z-resilience in an adaptive setting for the case of a CPA, gCCA2 and CCA2 attack respectively. Subsequent schemes build on the previous one, in a incremental way, so that it is possible to obtain increasing security at the cost of a slight efficiency loss.

Considering the subtlety of the arguments, our proofs follow the structural approach advocated in [8] defining a sequence of attack games $\mathbf{G}_0$, $\mathbf{G}_1$, ..., all operating over the same underlying probability space. Starting from the actual adversarial game $\mathbf{G}_0$, we incrementally make slight modifications to the behavior of the oracles, thus changing the way the adversary's view is computed, while maintaining the view's distributions indistinguishable among the games. While this structural approach takes more space to write down, it is much less error-prone and much more understandable than a slicker "direct argument" (e.g., compare [7] and [8]).

4.1 z-Resilience against CPA attack

As a warm-up before addressing the more challenging case of chosen ciphertext security, we describe a simpler CPA-secure scheme. Our scheme naturally builds upon previous works [19, 21], but achieves a much more appropriate notion of *adaptive* security, which those previous schemes do not enjoy.

The Key Generation Algorithm. The first step in the key generation algorithm $\mathsf{KeyGen}(1^\lambda, z)$ is to define a group $\mathbb{G}$ of order q, for a random λ-bit-long prime q such that $p = 2q + 1$ is also prime, in which the DDH assumption is believed to hold. This is accomplished selecting a random prime q with the above two properties and a random element g_1 of order q modulo p: the group $\mathbb{G}$ is then set to be the subgroup of $\mathbb{Z}_p^*$ generated by g_1, i.e. $\mathbb{G} = \{g_1^i \bmod p : i \in \mathbb{Z}_q\} \subset \mathbb{Z}_p^*$. A random $w \leftarrow_R \mathbb{Z}_q$ is then chosen and used to compute $g_2 = g_1^w$. (In what follows, all computations are mod q in the exponent, and mod p elsewhere.) Then, the key generation algorithm selects two random z-degree polynomials[3] $Z_1(\xi)$ and $Z_2(\xi)$ over $\mathbb{Z}_q$, and computes the values: $h_0 \doteq g_1^{Z_{1,0}} \cdot g_2^{Z_{2,0}}, \ldots, h_z \doteq g_1^{Z_{1,z}} \cdot g_2^{Z_{2,z}}$. Finally, the pair (PK, SK_{BE}) is given in output, where $PK \doteq \langle g_1, g_2, h_0, \ldots, h_z \rangle$ and $SK_{\mathsf{BE}} \doteq \langle Z_1, Z_2 \rangle$.

The Registration Algorithm. Each time a new user $i > z$ (in all our schemes, we reserve the first indices $0 \ldots z$ for "special purposes"), decides to subscribe to the system, the center provides him with a decoder box containing the secret key: $SK_i \doteq \langle i, Z_{1,i}, Z_{2,i} \rangle$.

The Encryption Algorithm. The encryption algorithm Enc is given in Figure 1. It receives as input the public key PK, a session key s and a set $\mathcal{R} = \{j_1, \ldots, j_z\}$ of revoked users and returns the enabling block $\mathcal{T}$. If there are less than z revoked users, the remaining indices are set to $1 \ldots (z - |\mathcal{R}|)$, which are never given to any "real" user.

The Decryption Algorithm. If a legitimate user i wants to recover the session key embedded in the enabling block $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}) \rangle$, he can proceed as in Figure 2. If i is a revoked user (i.e. $i \in \{j_1, \ldots, j_z\}$), the algorithm fails in step $D2$, since the interpolation points $j_1, \ldots, j_z, i$ are not pairwise distinct.

Security. As shown in the theorem below, the z-resilience of the above scheme relies on the Decisional Diffie-Hellman (DDH) assumption.

[3] For conciseness, we will use the following notation: $Z_{1,i} \doteq Z_1(i)$ and $Z_{2,i} \doteq Z_2(i)$.

$E1.$ $r_1 \leftarrow_R \mathbb{Z}_q$

$E2.$ $u_1 \leftarrow g_1^{r_1}$

$E3.$ $u_2 \leftarrow g_2^{r_1}$

$E4.$ $H_t \leftarrow h_t^{r_1}, \quad t = 0 \ldots z$

$E5.$ $H_{j_t} \leftarrow \mathsf{EXP\text{-}LI}(0, \ldots, z; H_0, \ldots, H_z)(j_t), \quad t = 1 \ldots z$

$E6.$ $S \leftarrow s \cdot H_0$

$E7.$ $\mathcal{T} \leftarrow \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}) \rangle$

Figure 1: Encryption algorithm: $\mathsf{Enc}(PK, s, \mathcal{R})$

$D1.$ $H_i \leftarrow u_1^{Z_{1,i}} \cdot u_2^{Z_{2,i}}$

$D2.$ $s \leftarrow S/\mathsf{EXP\text{-}LI}(j_1, \ldots, j_z, i; H_{j_1}, \ldots, H_{j_z}, H_i)(0)$

Figure 2: Decryption algorithm (for user i) $\mathsf{Dec}(i, \mathcal{T})$

Theorem 4 *If the* DDH *problem is hard in* $\mathbb{G}$*, then the above Broadcast Encryption Scheme is* z*-resilient against chosen plaintext attacks. In particular, for all probabilistic* poly*-time algorithm* $\mathcal{A}$*, we have that* $\mathsf{Adv}^{\mathsf{CPA}}_{\mathsf{BE},\mathcal{A}}(\lambda) \leq \nu(\lambda)$.

Proof: We define a sequence of "indistinguishable" games $\mathbf{G}_0, \ldots$, where $\mathbf{G}_0$ is the original game, and the last game clearly gives no advantage to the adversary.

Game $\mathbf{G}_0$. In game $\mathbf{G}_0$, $\mathcal{A}$ receives the public key PK and adaptively queries the corruption oracle $\mathsf{Cor}_{SK_{\mathsf{BE}}}(\cdot)$. Then, she queries the encryption oracle $\mathcal{E}_{PK,\mathcal{R},\sigma}(\cdot,\cdot)$ on (s_0, s_1), where $\mathcal{R}$ must contain all users that $\mathcal{A}$ compromised through the oracle $\mathsf{Cor}_{SK_{\mathsf{BE}}}(\cdot)$; $\mathcal{A}$ receives back the enabling block $\mathcal{T}^*$. At this point, $\mathcal{A}$ outputs her guess $\sigma^* \in \{0, 1\}$. Let T_0 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_0$.

Game $\mathbf{G}_1$. Game $\mathbf{G}_1$ is identical to game $\mathbf{G}_0$, except that, in game $\mathbf{G}_1$, step $E4$ of the encryption algorithm in Figure 1, is replaced with the following:

$$E4'. \quad H_t \leftarrow u_1^{Z_{1,t}} \cdot u_2^{Z_{2,t}}, \quad t = 0 \ldots z$$

By the properties of the Lagrange Interpolation in the Exponent, it is clear that step $E4'$ computes the same values H_t, $t = 0 \ldots z$ as step $E4$. The point of this change is just to make explicit any functional dependency of the above quantities on u_1 and u_2. Let T_1 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_1$; clearly, it holds that $\boxed{\Pr[T_0] = \Pr[T_1]}$.

Game $\mathbf{G}_2$. To turn game $\mathbf{G}_1$ into game $\mathbf{G}_2$ we make another change to the encryption oracle used in game $\mathbf{G}_1$. In game $\mathbf{G}_2$ steps $E1, E3$ are replaced with the following:

$$E1'. \quad r_1 \leftarrow_R \mathbb{Z}_q, \quad r_2 \leftarrow_R \mathbb{Z}_q \setminus \{r_1\}$$
$$E3'. \quad u_2 \leftarrow g_2^{r_2}$$

Let T_2 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_2$. Notice that while in game $\mathbf{G}_1$ the values u_1 and u_2 are obtained using the same value r_1, in game $\mathbf{G}_2$ they are independent subject to $r_1 \neq r_2$. Therefore, using a standard reduction argument, any non-negligible difference in behavior between $\mathbf{G}_1$ and $\mathbf{G}_2$ can be used to construct a PPT algorithm $\mathcal{A}_1$ that is able to distinguish Diffie-Hellman tuples from totally random tuples with non negligible advantage. Hence, $\boxed{\left|\Pr[T_2] - \Pr[T_1]\right| \leq \epsilon_1}$ for some negligible ϵ_1.

Game $\mathbf{G}_3$. To define game $\mathbf{G}_3$, we again modify the encryption oracle as follows:

$$E6'. \quad e \leftarrow_R \mathbb{Z}_q, \quad S \leftarrow g_1^e$$

Let T_3 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_3$. Because of this last change, the challenge no longer contains σ, nor does any other information in the adversary's view; therefore, we have that $\boxed{\Pr[T_3] = \frac{1}{2}}$. Moreover, we can prove (see Appendix 4.3, Lemma 9), that the adversary has the same chances to guess σ in both game $\mathbf{G}_2$ and $\mathbf{G}_3$, i.e. $\boxed{\Pr[T_3] = \Pr[T_2]}$.

Finally, combining all the intermediate results together, we can conclude that adversary $\mathcal{A}$'s advantage is negligible; more precisely: $\mathsf{Adv}^{\mathsf{CPA}}_{\mathsf{BE},\mathcal{A}}(\lambda) \leq \epsilon_1$. □

A comparison with the CPA schemes of [19, 21]. Our CPA scheme extends those proposed in [19, 21] by using two generators. This improvement turns out to be crucial: the adaptive security of our CPA scheme hinges heavily upon this change. In particular, the presence of two generators plays a key role in the reduction from the DDH problem mentioned in the description of game $\mathbf{G}_2$ in Theorem 4. More specifically, when setting up the simulation of the adaptive attack scenario (defined in Section 3.1), the two distinct generators used in the public key of the scheme provide the perfect place where to embed the first two elements g_1 and g_2 of the DDH "challenge" at hand. Doing so, the simulator can choose the rest of the public key in a "honest" way, and hence it will know all the corresponding secret keys (i.e. the polynomials $Z_1(\xi)$ and $Z_2(\xi)$). This in turn allows the simulator to answer any corruption query that the adversary may want to carry out before querying the encryption oracle.
On the contrary, the use of a single generator in both CPA schemes of [19, 21] leads to a reduction in which the simulator does not know the entire secret key (in particular, the constant term of the secret polynomial is unknown to the simulator; cfr. Theorem 1 of [21]). As a consequence, there seems to be no way to answer corruption queries, so that the adaptive attack scenario from Section 3.1 cannot be properly simulated: thus the reduction argument does not go through.

4.2 z-Resilience against gCCA2 Attack

Once we have constructed a Broadcast Encryption Scheme z-resilient against CPA attacks, it is natural to try to devise an extension achieving adaptive chosen ciphertext security. This was already attempted by [21], but they do not elaborate (neither formally nor informally) on what an "adaptive chosen ciphertext attack" on a Broadcast Encryption Scheme exactly is. As a consequence, in Theorem 3 of [21], the authors only show the security of their scheme against an adversary that does not participate in the system, whereas (as we will argue at the end of this section) their scheme is certainly *not* CCA2-secure with respect to even a *single* malicious revoked user.

To achieve CCA2 security, we will first try to apply the standard technique of [7, 8] to the scheme presented in Section 4.1. Unfortunately, this natural approach does *not* completely solve the CCA2 problem; still, it leads us to an interesting scheme that achieves the (sligthly weaker) notion of generalized chosen ciphertext security.

The Key Generation Algorithm. As before, the first task of the key generation algorithm is to select a random group $\mathbb{G} \subset \mathbb{Z}_p^*$ of prime order q and two random generators $g_1, g_2 \in \mathbb{G}$. Then, KeyGen selects six random z-degree polynomials[4] $X_1(\xi)$, $X_2(\xi)$, $Y_1(\xi)$, $Y_2(\xi)$, $Z_1(\xi)$ and $Z_2(\xi)$ over $\mathbb{Z}_q$, and computes the values $c_t \doteq g_1^{X_{1,t}} \cdot g_2^{X_{2,t}}$, $d_t \doteq g_1^{Y_{1,t}} \cdot g_2^{Y_{2,t}}$ and $h_t \doteq g_1^{Z_{1,t}} \cdot g_2^{Z_{2,t}}$, for $t = 0 \dots z$.

[4]For conciseness, we will use the following notation: $X_{1,i} \doteq X_1(i), X_{2,i} \doteq X_2(i), Y_{1,i} \doteq Y_1(i), Y_{2,i} \doteq Y_2(i), Z_{1,i} \doteq Z_1(i)$ and $Z_{2,i} \doteq Z_2(i)$.

Finally, KeyGen chooses at random a hash function $\mathcal{H}$ from a family $\mathcal{F}$ of collision resistant hash functions,[5] and outputs the pair (PK, SK_{BE}), where $PK \doteq \langle g_1, g_2, c_0, \ldots, c_z, d_0, \ldots, d_z, h_0, \ldots, h_z, \mathcal{H}\rangle$ and $SK_{\mathsf{BE}} \doteq \langle X_1, X_2, Y_1, Y_2, Z_1, Z_2\rangle$.

The Registration Algorithm. Each time a new user $i > z$ subscribes to the system, the center provides him with a decoder box containing the secret key $SK_i \doteq \langle i, X_{1,i}, X_{2,i}, Y_{1,i}, Y_{2,i}, Z_{1,i}, Z_{2,i}\rangle$.

The Encryption Algorithm. Using the idea of [7, 8], in order to obtain non-malleable ciphertexts, we "tag" each encrypted message so that it can be verified before proceeding with the actual decryption. In the broadcast encryption scenario, where each user has a different decryption key, the tag cannot be a single point — we need to distribute an entire EXP-polynomial $\mathcal{V}(x)$. This is accomplished appending $z+1$ tags to the ciphertext: each user i first computes the tag v_i using his private key and then verifies the validity of the ciphertext by checking the interpolation of the $z+1$ values in point i against its v_i.

The encryption algorithm Enc receives as input the public key PK, the session key s to be embedded within the enabling block and a set $\mathcal{R} = \{j_1, \ldots, j_z\}$ of revoked users. It proceeds as described in Figure 3, and finally it outputs $\mathcal{T}$.

$E1.$ $r_1 \leftarrow_R \mathbb{Z}_q$
$E2.$ $u_1 \leftarrow g_1^{r_1}$
$E3.$ $u_2 \leftarrow g_2^{r_1}$
$E4.$ $H_t \leftarrow h_t^{r_1}, \quad t = 0 \ldots z$
$E5.$ $H_{j_t} \leftarrow \mathsf{EXP\text{-}LI}(0, \ldots, z; H_0, \ldots, H_z)(j_t), \quad t = 1 \ldots z$
$E6.$ $S \leftarrow s \cdot H_0$
$E7.$ $\alpha \leftarrow \mathcal{H}(S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}))$
$E8.$ $v_t \leftarrow c_t^{r_1} \cdot d_t^{r_1 \alpha}, \quad t = 0 \ldots z$
$E9.$ $\mathcal{T} \leftarrow \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v_0, \ldots, v_z\rangle$

Figure 3: Encryption algorithm $\mathsf{Enc}(PK, s, \mathcal{R})$

The Decryption Algorithm. If a legitimate user i wants to recover the session key embedded in the enabling block $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v_0, \ldots, v_z\rangle$, he can proceed as in Figure 4. If i is a revoked user, the algorithm fails in step $D6$, since the interpolation points $j_1, \ldots, j_z, i$ are not pairwise distinct.

Security. As mentioned above, the presence of many decryption keys leads to the use of an EXP-polynomial $\mathcal{V}(x)$ to tag the encryption of the message. This in turn makes the ciphertext malleable: since each user i can verify the value of $\mathcal{V}(x)$ only in one point, the adversary can modify the v_j's values and construct a different EXP-polynomial $\mathcal{V}'(x)$ intersecting $\mathcal{V}(x)$ at point i — thus fooling user i to accept as valid a corrupted ciphertext. In the next section we show a non-trivial solution to this problem; here, we assess the z-resilience of the Broadcast Encryption Scheme presented above against a gCCA2 attack. As already discussed in Section 3.1, to this aim it is necessary to introduce a family of equivalence relations $\{\Re_i\}$: intuitively, two ciphertexts $\mathcal{T}$ and $\mathcal{T}'$ are equivalent for user i if they have the same "data" components, and the tag "relevant to user i" is correctly verified, i.e. $v_i = v'_i$ (even though other "irrelevant" tags could be different). Clearly, this relation is efficiently computable and i-decryption-respecting.

[5] Recall, it is hard to find $x \neq y$ such that $\mathcal{H}(x) = \mathcal{H}(y)$ for a random member $\mathcal{H}$ of $\mathcal{F}$.

$D1.$ $\alpha \leftarrow \mathcal{H}(S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}))$

$D2.$ $\bar{v}_i \leftarrow u_1^{X_{1,i}+Y_{1,i}\alpha} \cdot u_2^{X_{2,i}+Y_{2,i}\alpha}$

$D3.$ $v_i \leftarrow \mathsf{EXP\text{-}LI}(0, \ldots, z; v_0, \ldots, v_z)(i)$

$D4.$ **if** $v_i = \bar{v}_i$

$D5.$ **then** $H_i \leftarrow u_1^{Z_{1,i}} \cdot u_2^{Z_{2,i}}$

$D6.$ $s \leftarrow S/\mathsf{EXP\text{-}LI}(j_1, \ldots, j_z, i; H_{j_1}, \ldots, H_{j_z}, H_i)(0)$

$D7.$ **return** s

$D8.$ **else return** $\perp$

Figure 4: Decryption algorithm (for user i) $\mathsf{Dec}(i, \mathcal{T})$

Definition 5 (Equivalence Relation)
Consider $\mathcal{V}(x) = \mathsf{EXP\text{-}LI}(0, \ldots, z; v_0, \ldots, v_z)(x)$ *and* $\mathcal{V}'(x) = \mathsf{EXP\text{-}LI}(0, \ldots, z; v'_0, \ldots, v'_z)(x)$. *Given a user* i, *and the two enabling blocks* $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v_0, \ldots, v_z \rangle$ *and* $\mathcal{T}' = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v'_0, \ldots, v'_z \rangle$, *we say that* $\mathcal{T}$ *is* equivalent *to* $\mathcal{T}'$ *with respect to user* i, *and we write* $\Re_i(\mathcal{T}, \mathcal{T}')$, *if the two* EXP*-polynomials* $\mathcal{V}(x)$ *and* $\mathcal{V}'(x)$ *intersect at point* i, *i.e.* $v_i = \mathcal{V}(i) = \mathcal{V}'(i) = v'_i$.

Theorem 6 *If the* DDH *Problem is hard in* $\mathbb{G}$ *and* $\mathcal{H}$ *is chosen from a collision-resistant hash functions family* $\mathcal{F}$, *then the above Broadcast Encryption Scheme is* z*-resilient against generalized chosen ciphertext attacks, under the family of equivalence relations* $\{\Re_i\}$.

Proof: To prove this theorem, we pursue the same approach as in the proof of Theorem 4, where the starting scenario of the sequence of games is defined as in the definition of the adaptive gCCA2 attack.

Game $\mathbf{G}_0$. Recall that in game $\mathbf{G}_0$, $\mathcal{A}$ receives the public key PK and adaptively interleaves queries to the corruption oracle $\mathsf{Cor}_{SK_{\mathsf{BE}}}(\cdot)$ with queries to the decryption oracle $\mathcal{D}_{SK_{\mathsf{BE}}}(\cdot, \cdot)$. Then, she queries the encryption oracle $\mathcal{E}_{PK,\mathcal{R},\sigma}(\cdot, \cdot)$ on (s_0, s_1), where $\mathcal{R}$ must contain all users that $\mathcal{A}$ compromised through the oracle $\mathsf{Cor}_{SK_{\mathsf{BE}}}(\cdot)$; $\mathcal{A}$ receives back the enabling block $\mathcal{T}^*$. Then, $\mathcal{A}$ can again query the decryption oracle $\mathcal{D}_{SK_{\mathsf{BE}}}(i, \mathcal{T})$, restricted only in that $\neg\Re_i(\mathcal{T}, \mathcal{T}^*)$. Finally, she outputs her guess $\sigma^* \in \{0, 1\}$. Let T_0 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_0$.

Game $\mathbf{G}_1$. Game $\mathbf{G}_1$ is identical to game $\mathbf{G}_0$, except that, in game $\mathbf{G}_1$, steps $E4, E8$ of the encryption algorithm in Figure 3, are replaced with the following:

$$E4'. \quad H_t \leftarrow u_1^{Z_{1,t}} \cdot u_2^{Z_{2,t}}, \quad t = 0 \ldots z$$

$$E8'. \quad v_t \leftarrow u_1^{X_{1,t}+Y_{1,t}\alpha} \cdot u_2^{X_{2,t}+Y_{2,t}\alpha} \quad t = 0 \ldots z$$

By the properties of the Lagrange Interpolation in the Exponent, it is clear that step $E4'$ computes the same values H_{j_t}, $t = 0 \ldots z$ as steps $E4$; similarly, step $E8'$ computes the same values v_t, $t = 0 \ldots z$ as step $E8$. The point of these changes is just to make explicit any functional dependency of the above quantities on u_1 and u_2.

Let T_1 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_1$. Clearly, it holds that $\boxed{\Pr[T_0] = \Pr[T_1]}$.

Game $\mathbf{G}_2$. To turn game $\mathbf{G}_1$ into game $\mathbf{G}_2$ we make another change to the encryption oracle used

in game $\mathbf{G}_1$. In game $\mathbf{G}_2$ steps $E1, E3$ are replaced with the following:

$$E1'. \quad r_1 \leftarrow_R \mathbb{Z}_q, \quad r_2 \leftarrow_R \mathbb{Z}_q \setminus \{r_1\}$$
$$E3'. \quad u_2 \leftarrow g_2^{r_2}$$

Let T_2 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_2$. Notice that while in game $\mathbf{G}_1$ the values u_1 and u_2 are obtained using the same value r_1, in game $\mathbf{G}_2$ they are independent subject to $r_1 \neq r_2$. Therefore, using a standard reduction argument, any non-negligible difference in behavior between $\mathbf{G}_1$ and $\mathbf{G}_2$ can be used to construct a PPT algorithm $\mathcal{A}_1$ that is able to distinguish Diffie-Hellman tuples from totally random tuples with non negligible advantage. Hence, $\boxed{\left|\Pr[T_2] - \Pr[T_1]\right| \leq \epsilon_1}$ for some negligible ϵ_1.

Game $\mathbf{G}_3$. To define game $\mathbf{G}_3$ we slightly modify the decryption oracle: instead of using the algorithm in Figure 4, in game $\mathbf{G}_3$ steps $D2, D4, D5$ are replaced with the following:

$$D2'. \quad \bar{v}_i \leftarrow u_1^{(X_{1,i}+Y_{1,i}\alpha)+(X_{2,i}+Y_{2,i}\alpha)\cdot w}$$
$$D4'. \quad \textbf{if } (u_2 = u_1^w \wedge v_i = \bar{v}_i)$$
$$D5'. \quad \textbf{then } H_i \leftarrow u_1^{Z_{1,i}+Z_{1,i}\cdot w}$$

The rationale behind these changes is that we want to strengthen the condition that the enabling block has to meet in order to be considered valid and hence to be decrypted. This will make it easier to show the security of the scheme; however, for these changes to be useful, there should be no observable difference in the way invalid enabling blocks are "caught" in games $\mathbf{G}_2$ and $\mathbf{G}_3$. To make it formal, we now introduce the following two events: let T_3 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_3$, and let R_3 be the event that $\mathcal{A}$ submits some decryption query that would have been decrypted in game $\mathbf{G}_2$ but is rejected in game $\mathbf{G}_3$; in other words, R_3 is the event that some decryption query that would have passed the test in step $D4$ of the decryption oracle used in game $\mathbf{G}_2$, fails to pass the test in step $D4'$ used in game $\mathbf{G}_3$. Clearly, $\mathbf{G}_2$ and $\mathbf{G}_3$ are identical until event R_3 occurs; hence, if R_3 never occurs, the adversary has the same chances to win in both the two games, i.e. (using Lemma 1) $T_3 \wedge \neg R_3 \equiv T_2 \wedge \neg R_3 \Rightarrow \boxed{\left|\Pr[T_3] - \Pr[T_2]\right| \leq \Pr[R_3]}$.

To bound the last probability, we consider two more games, $\mathbf{G}_4$ and $\mathbf{G}_5$.

Game $\mathbf{G}_4$. To define game $\mathbf{G}_4$, we again modify the encryption oracle as follows:

$$E6'. \quad e \leftarrow_R \mathbb{Z}_q, \quad S \leftarrow g_1^e$$

Let T_4 be the event that $\sigma = \sigma^*$ in game $\mathbf{G}_4$. Because of this last change, the challenge no longer contains the bit σ, nor does any other information in the adversary's view; therefore, we have that $\boxed{\Pr[T_4] = \frac{1}{2}}$.

Let R_4 be the event that $\mathcal{A}$ submits some decryption query that would have been decrypted in game $\mathbf{G}_2$ but is rejected in game $\mathbf{G}_4$; in other words, R_4 is the event that some decryption query that would have passed the test in step $D4$ of the decryption oracle used in game $\mathbf{G}_2$, fails to pass the test in step $D4'$ used in game $\mathbf{G}_4$. In Appendix 4.3, we prove (Lemma 10) that those events happen with the same probability as the corresponding events of game $\mathbf{G}_3$, i.e. $\boxed{\Pr[T_4] = \Pr[T_3]}$ and $\boxed{\Pr[R_4] = \Pr[R_3]}$.

Game $\mathbf{G}_5$. In this game, we again modify the decryption algorithm, adding the following *special rejection rule*, whose goal is to prevent the adversary from submitting illegal enabling blocks to the decryption oracle, once she has received her challenge.

After $\mathcal{A}$ receives her challenge $\mathcal{T}^* = \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*}), v_0^*, \ldots, v_z^* \rangle$, the decryption oracle rejects any query $\langle i, \mathcal{T} \rangle$, with $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v_0, \ldots, v_z \rangle$ such that $\langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}) \rangle \neq \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*}) \rangle$, but $\alpha = \alpha^*$, and it does so before executing the test in step $D4'$.

Notice that in the gCCA2 setting the adversary is not allowed to query the decryption oracle $\mathsf{Dec}(i, \mathcal{T})$ on enabling blocks $\Re_i$-equivalent to the challenge $\mathcal{T}^*$. Therefore, when the *special rejection rule* is applied, we already know that it holds $\neg\Re_i(\mathcal{T}, \mathcal{T}^*)$.

Let C_5 be the event that the adversary submits a decryption query that is rejected using the above *special rejection rule*; let R_5 be the event that $\mathcal{A}$ submits some decryption query that would have passed the test in step $D4$ of the decryption oracle used in game $\mathbf{G}_2$, but fails to pass the test in step $D4'$ used in game $\mathbf{G}_5$. Notice that this implies that such a query passed the $\Re_i$-equivalence test and the *special rejection rule*, because otherwise step $D4'$ wouldn't have been executed. Clearly, $\mathbf{G}_4$ and $\mathbf{G}_5$ are identical until event C_5 occurs, i.e. $R_5 \wedge \neg C_5 \equiv R_4 \wedge \neg C_5 \Rightarrow \boxed{\big|\Pr[R_5] - \Pr[R_4]\big| \leq \Pr[C_5]}$, where the implication follows from Lemma 1.

Our final task is to show that events C_5 and R_5 occur with negligible probability: while the argument to bound event C_5 is based on the collision resistance assumption for the family $\mathcal{F}$ (using a standard reduction argument, we can construct a PPT algorithm $\mathcal{A}_2$ that breaks the collision resistance assumption with non negligible advantage), the argument to bound event R_5 hinges upon the fact that the adversary is not allowed to submit queries that are "$\Re_i$-related" to her challenge, and upon information-theoretic considerations (as proven in Appendix 4.3, Lemma 11). From these considerations, we obtain that $\boxed{\Pr[C_5] \leq \epsilon_2}$ and $\boxed{\Pr[R_5] \leq \frac{Q_{\mathcal{A}}(\lambda)}{q}}$, where ϵ_2 is a negligible quantity and $Q_{\mathcal{A}}(\lambda)$ is an upper bound on the number of decryption queries made by the adversary.

Finally, combining the intermediate results, we can conclude that adversary $\mathcal{A}$'s advantage is negligible; more precisely: $\mathsf{Adv}^{\mathsf{gCCA2}}_{\mathsf{BE},\mathcal{A}}(\lambda) \leq \epsilon_1 + \epsilon_2 + Q_{\mathcal{A}}(\lambda)/q$. □

A comparison with the CCA2 attempt of [21]. As already noticed when describing the encryption algorithm, in the broadcast encryption setting it is not safe to use a single point as validating tag; it is for that reason that in our gCCA2 solution we use a validating EXP-polynomial $\mathcal{V}(x)$. On the contrary, in [21] the authors didn't recognize the inherent insecurity of using a single tag and proposed a scheme that not only uses a single generator, but also tags ciphertexts with just one point. Consequently, the information distributed to the users of the system to enable them to check the validity of a given ciphertext (namely, x_1, x_2, y_1, y_2 in the notation of [21]) is the *same* for all participants. To verify the validity of a ciphertext $\mathcal{T}$, a user recomputes the tag $\bar{v} = F_a^{x_1 + y_1\alpha} \cdot F_b^{x_2 + y_2\alpha}$ from quantities present in the ciphertext itself and from the secret information x_1, x_2, y_1, y_2 (common to all users). The value $\bar{v}$ is then compared against the tag v in $\mathcal{T}$.

This means that revoking a user does not affect his/her ability to check the validity of a ciphertext; furthermore, validating a ciphertext is effectively equivalent to computing the corresponding tag. This implies that any revoked user is able to construct new, legal ciphertexts from any encrypted message; in other words, ciphertexts are *malleable* and hence the scheme cannot be CCA2 secure. More precisely, even an adversary that non-adaptively corrupts just a single user, can break the scheme with a single decryption query: upon receiving the challenge ciphertext from the encryption oracle, the adversary changes it in some easily-reversible way, computes the proper tag (exploiting the knowledge of x_1, x_2, y_1, y_2 that she got from the revoked user) and asks the decryption oracle to decrypt such modified ciphertext. Once the adversary gets the decrypted message back, she can easily tell which message was hidden within her challenge, breaking the scheme.

4.3 z-Resilience against CCA2 Attack

In Section 4.2, we saw how a direct application of the standard technique of [7, 8] does not provide a complete solution to the CCA2 problem, but only suffices for gCCA2 security. As proven in Lemma 11 (see Appendix 4.3), the restriction imposed by the gCCA2 attack (namely, forbidding the adversary to submit decryption queries $\langle i, \mathcal{T} \rangle$ such that $\Re_i(\mathcal{T}, \mathcal{T}^*)$ holds) is essential for the security of the previous Broadcast Encryption Scheme. Indeed, given a challenge $\mathcal{T}^*$ with tag sequence $v_0 \dots v_z$, it is trivial to come up with a different sequence $v'_0 \dots v'_z$ such that $v_i = v'_i$, resulting in a "different" enabling block $\mathcal{T}' \neq \mathcal{T}^*$: however, $\mathsf{Dec}(i, \mathcal{T}^*) = \mathsf{Dec}(i, \mathcal{T}')$, allowing the adversary to "break" the CCA2 security.

Although we feel that gCCA2 security is enough for most applications of Broadcast Encryption Schemes, it is possible to non-trivially modify the Broadcast Encryption Scheme presented in Section 4.2 to obtain CCA2 security (with only a slight efficiency loss). The modified scheme, presented in this section, maintains the same Key Generation and Registration algorithms described before; the essential modifications involve the operations used to construct the enabling block. In particular, to achieve CCA2 security, it is necessary to come up with some trick to make the tag sequence $v_0, \dots, v_z$ non-malleable. To this aim, we will use any secure (deterministic) *message authentication code* (MAC) to guarantee the integrity of the entire sequence. In fact, we only need any *one-time* MAC, satisfying the following simple property: given a (unique) correct value $\mathsf{MAC}_k(M)$ for some message M (under key k), it is infeasible to come up with a correct (unique) value of $\mathsf{MAC}_k(M')$, for any $M' \neq M$.

The Encryption Algorithm. The encryption algorithm Enc receives as input the public key PK, the session key s to be embedded within the enabling block and a set $\mathcal{R} = \{j_1, \dots, j_z\}$ of revoked users. To construct the enabling block $\mathcal{T}$, the encryption algorithm (defined in Figure 3) operates similarly to the gCCA2 encryption algorithm: the main difference is that now a MAC key k, randomly chosen from the MAC key space $\mathcal{K}$, is used to MAC the tag sequence $v_0, \dots, v_z$, and is encapsulated within $\mathcal{T}$ along with the session key s.

$E1.$ $r_1 \leftarrow_R \mathbb{Z}_q$
$E2.$ $u_1 \leftarrow g_1^{r_1}$
$E3.$ $u_2 \leftarrow g_2^{r_1}$
$E4.$ $H_t \leftarrow h_t^{r_1}, \quad t = 0 \dots z$
$E5.$ $H_{j_t} \leftarrow \mathsf{EXP\text{-}LI}(0, \dots, z; H_0, \dots, H_z)(j_t), \quad t = 1 \dots z$
$E6.$ $k \leftarrow_R \mathcal{K}$
$E7.$ $S \leftarrow (s \parallel k) \cdot H_0$
$E8.$ $\alpha \leftarrow \mathcal{H}(S, u_1, u_2, (j_1, H_{j_1}), \dots, (j_z, H_{j_z}))$
$E9.$ $v_t \leftarrow c_t^{r_1} \cdot d_t^{r_1 \alpha}, \quad t = 0 \dots z$
$E10.$ $\tau \leftarrow \mathsf{MAC}_k(v_0, \dots, v_z)$
$E11.$ $\mathcal{T} \leftarrow \langle S, u_1, u_2, (j_1, H_{j_1}), \dots, (j_z, H_{j_z}), v_0, \dots, v_z, \tau \rangle$

Figure 5: Encryption algorithm $\mathsf{Enc}(PK, s, \mathcal{R})$

The Decryption Algorithm. If a legitimate user i wants to recover the session key embedded in the enabling block $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \dots, (j_z, H_{j_z}) \rangle$, he can proceed as in Figure 6. If i is a revoked user, the algorithm fails in step $D6$, since the interpolation points $j_1, \dots, j_z, i$ are not pairwise distinct.

$D1.$ $\alpha \leftarrow \mathcal{H}(S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}))$

$D2.$ $\bar{v}_i \leftarrow u_1^{X_{1,i}+Y_{1,i}\alpha} \cdot u_2^{X_{2,i}+Y_{2,i}\alpha}$

$D3.$ $v_i \leftarrow \mathsf{EXP\text{-}LI}(0, \ldots, z; v_0, \ldots, v_z)(i)$

$D4.$ **if** $v_i = \bar{v}_i$

$D5.$ **then** $H_i \leftarrow u_1^{Z_{1,i}} \cdot u_2^{Z_{2,i}}$

$D6.$ $\quad s \parallel k \leftarrow S/\mathsf{EXP\text{-}LI}(j_1, \ldots, j_z, i; H_{j_1}, \ldots, H_{j_z}, H_i)(0)$

$D7.$ $\quad$ extract s and k from $s \parallel k$

$D8.$ $\quad$ **if** $\tau \neq \mathsf{MAC}_k(v_0, \ldots, v_z)$

$D9.$ $\quad$ **then return** $\perp$

$D10.$ $\quad$ **else return** s

$D11.$ **else return** $\perp$

Figure 6: Decryption algorithm (for user i) $\mathsf{Dec}(i, \mathcal{T})$

SECURITY. The security analysis for this scheme is very subtle, because there is the risk of circularity in the use of the MAC key k. Namely, k is part of the ciphertext (since it is encapsulated, along with the session key s, within S); this means that α, the hash of the ciphertext, depends on k (at least Information-Theoretically), and thus the sequence of tags depends on k. In other words, we are MAC-ing something that depends on the MAC key k, which could be a problem. Luckily, the Information-Theoretic nature of the structural approach to the security analysis that we are pursuing (following [8]) allows us to prove that actually k is completely hidden within S, so that MAC-ing the resulting tag with k is still secure.

The solution to the CCA2 problem for Broadcast Encryption Schemes and the relative security analysis can be viewed as the main technical contribution of this paper; at the same time, the capability to resolve the apparent circularity in the use of the MAC demonstrates the importance of providing a formal model and precise definitions, without which it would have been much harder to devise a correct proof of security for the above scheme.

Theorem 7 *If the* DDH *Problem is hard in* $\mathbb{G}$*,* $\mathcal{H}$ *is chosen from a collision-resistant hash functions family* $\mathcal{F}$ *and* MAC *is a one-time message authentication code, then the above Broadcast Encryption Scheme is z-resilient against chosen ciphertext attacks.*

Proof: The proof proceeds defining a sequence of games similar to that presented in Theorem 6. The definition of games $\mathbf{G}_0, \ldots, \mathbf{G}_5$ closely follow the exposition given in Theorem 6: however, the statements of all lemmas (and their proofs) need to be changed to accommodate for the use of the MAC. In particular, we can easily state and prove a lemma analogous to Lemma 10, where the only difference is the presence of information about the MAC key k in the challenge (see Lemma 12). More importantly, to bound the probability $\Pr[R_5]$ we introduce a new game $\mathbf{G}_6$ to deal with the use of the MAC in the enabling block, while a lemma similar to Lemma 11 is used to bound the probability of event R_6 defined in game $\mathbf{G}_6$ (see Appendix 4.3 for the details on the proofs).

Game $\mathbf{G}_6$. To define this game, we modify the decryption algorithm, adding the following *second special rejection rule*, whose goal is to detect illegal enabling blocks submitted by the adversary to the decryption oracle, once she has received her challenge. Notice that, while the *special rejection rule*, defined in game $\mathbf{G}_5$, is used to reject adversary's queries aiming at exploiting any weakness in

the collision-resistant hash family $\mathcal{F}$, the *second special rejection rule* is used to reject ciphertexts aiming at exploiting any weakness in the MAC scheme.

> After $\mathcal{A}$ receives her challenge $\mathcal{T}^* = \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*}), v_0^*, \ldots, v_z^*, \tau^* \rangle$, the decryption oracle rejects any query $\langle i, \mathcal{T} \rangle$, with $\mathcal{T} = \langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}), v_0, \ldots, v_z, \tau \rangle$ such that $\langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}) \rangle = \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*}) \rangle$ and $(v_0, \ldots, v_z) \neq (v_0^*, \ldots, v_z^*)$, but $\tau = \mathsf{MAC}_{k^*}(v_0, \ldots, v_z)$, and it does so before executing the test in step $D4'$, and before applying the *special rejection rule*.

Let M_6 be the event that the adversary submits a decryption query that is rejected in game $\mathbf{G}_6$ using the *second special rejection rule*; let C_6 be the event that the adversary submits a decryption query that is rejected in game $\mathbf{G}_6$ using the *special rejection rule*; let R_6 be the event that $\mathcal{A}$ submits some decryption query that would have passed both the test in step $D4$ and in step $D8$ of the decryption oracle used in game $\mathbf{G}_2$, but fails to pass the test in step $D4'$ used in game $\mathbf{G}_6$. Notice that this implies that such a query passed both the *second special rejection rule* and the *special rejection rule*, because otherwise step $D4'$ wouldn't have been executed at all.

Event M_6 is closely related to the security of the one time MAC used in the scheme; in particular, any difference in behavior between game $\mathbf{G}_5$ and game $\mathbf{G}_6$ can be used to construct a PPT algorithm $\mathcal{A}_3$ that is able to forge a legal authentication code under a one-message attack with non-negligible probability, thus breaking the MAC scheme. Hence, $\boxed{\Pr[M_6] \leq \epsilon_3}$, for some negligible ϵ_3.

Moreover, since $\mathbf{G}_5$ and $\mathbf{G}_6$ are identical until event M_6 occurs, if it doesn't occur at all, they will proceed identically; i.e., by Lemma 1:

$$C_6 \wedge \neg M_6 \equiv C_5 \wedge \neg M_6 \Rightarrow \boxed{\big|\Pr[C_6] - \Pr[C_5]\big| \leq \Pr[M_6]}$$

$$R_6 \wedge \neg M_6 \equiv R_5 \wedge \neg M_6 \Rightarrow \boxed{\big|\Pr[R_6] - \Pr[R_5]\big| \leq \Pr[M_6]}.$$

Our final task is to bound the probability that events C_6 and R_6 occur: the argument to bound $\Pr[C_6]$ is based on the collision resistance assumption for the family $\mathcal{F}$, while the argument to bound $\Pr[R_6]$ hinges upon information-theoretic considerations (as proven in Appendix 4.3, Lemma 13). From those facts, we obtain that $\boxed{\Pr[C_6] \leq \epsilon_2}$ and $\boxed{\Pr[R_6] \leq \frac{Q_{\mathcal{A}}(\lambda)}{q}}$, where ϵ_2 is a negligible quantity and $Q_{\mathcal{A}}(\lambda)$ is an upper bound on the number of decryption queries made by the adversary.

Finally, combining the intermediate results, we can conclude that adversary $\mathcal{A}$'s advantage is negligible; more precisely: $\mathsf{Adv}^{\mathsf{CCA2}}_{\mathsf{BE},\mathcal{A}}(\lambda) \leq \epsilon_1 + \epsilon_2 + 2\epsilon_3 + Q_{\mathcal{A}}(\lambda)/q$. □

Acknowledgments

We wish to thank Jonathan Katz, Yevgeniy Kushnir, Antonio Nicolosi and Victor Shoup for helpful observations on an preliminary version of the paper and the anonymous referees for useful comments.

References

[1] J.H. An, Y. Dodis, and T. Rabin. On the Security of Joint Signature and Encryption. In *Advances in Cryptology - EuroCrypt '02*, pages 83–107. Springer-Verlag, 2002. LNCS 2332.

[2] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science - FOCS '97*, pages 394–403, 1997.

[3] D. Boneh. The Decision Diffie-Hellman Problem. In *Algorithmic Number Theory - ANTS-III*, pages 48–63. Springer-Verlag, 1998. LNCS 1423.

[4] R. Canetti, J. Garay, G. Itkis, D. Micciancio, M. Naor, and B. Pinkas. Multicast Security: A Taxonomy and some Efficient Constructions. In *Proceedings of IEEE INFOCOM '99*, volume 2, pages 708–716, 1999.

[5] R. Canetti, T. Malkin, and K. Nissim. Efficient Communication-Storage Tradeoffs for Multicast Encryption. In *Theory and Application of Cryptographic Techniques*, pages 459–474, 1999.

[6] B. Chor, A. Fiat, and N. Naor. Tracing Traitors. In *Advances in Cryptology - Crypto '94*, pages 257–270. Springer-Verlag, 1994. LNCS 839.

[7] R. Cramer and V. Shoup. A Practical Public Key Cryptosystem Provably Secure Against Adaptive Chosen Ciphertext Attack. In *Advances in Cryptology - Crypto '98*, pages 13–25. Springer-Verlag, 1998. LNCS 1462.

[8] R. Cramer and V. Shoup. Design and Analysis of Practical Public-Key Encryption Scheme Secure against Adaptive Chosen Ciphertext Attack. Manuscript, 2001.

[9] Y. Dodis and N. Fazio. Public Key Broadcast Encryption for Statless Receivers. In *Digital Rights Management - DRM '02*, 2002.

[10] Y. Dodis and N. Fazio. Public Key Trace and Revoke Scheme Secure against Adaptive Chosen Ciphertext Attack. In *Public Key cryptography - PKC '03*, pages 100–115. Springer-Verlag, 2003. LNCS 2567.

[11] A. Fiat and M. Naor. Broadcast Encryption. In *Advances in Cryptology - Crypto '93*, pages 480–491. Springer-Verlag, 1993. LNCS 773.

[12] E. Gafni, J. Staddon, and Y. L. Yin. Efficient Methods for Integrating Traceability and Broadcast Encryption. In *Advances in Cryptology - Crypto '99*, pages 372–387. Springer-Verlag, 1999. LNCS 1666.

[13] A Garay, J. Staddon, and A. Wool. Long-Lived Broadcast Encryption. In *Advances in Cryptology - Crypto 2000*, pages 333–352. Springer-Verlag, 2000. LNCS 1880.

[14] D. Halevy and A. Shamir. The LSD Broadcast Encryption Scheme. In *Advances in Cryptology - Crypto '02*, pages 47–60. Springer-Verlag, 2002. LNCS 2442.

[15] R. Kumar, S. Rajagopalan, and A. Sahai. Coding Constructions for Blacklisting Problems without Computational Assumptions. In *Advances in Cryptology - Crypto '99*, pages 609–623. Springer-Verlag, 1999. LNCS 1666.

[16] M. Luby and J. Staddon. Combinatorial Bounds for Broadcast Encryption. In *Advances in Cryptology - EuroCrypt '98*, pages 512–526. Springer-Verlag, 1998. LNCS 1403.

[17] D.A. McGrew and A.T. Sherman. Key Establishment in Large Dynamic Groups Using One-Way Function Trees. Manuscript, 1998.

[18] D. Naor, M. Naor, and J. Lotspiech. Revocation and Tracing Schemes for Stateless Receivers. In *Advances in Cryptology - Crypto '01*, pages 41–62. Springer-Verlag, 2001. LNCS 2139.

[19] M. Naor and B. Pinkas. Efficient Trace and Revoke Schemes. In *Financial Cryptography - FC 2000*, pages 1–20. Springer-Verlag, 2000. LNCS 1962.

[20] V. Shoup. A Proposal for an ISO Standard for Public-Key Encryption. Manuscript, 2001.

[21] W.G. Tzeng and Z.J. Tzeng. A Public-Key Traitor Tracing Scheme with Revocation Using Dynamics Shares. In *Public Key Cryptography - PKC '01*, pages 207–224. Springer-Verlag, 2001. LNCS 1992.

[22] D. Wallner, E. Harder, and R. Agee. Key Management for Multicast: Issues and Architectures. Available at `ftp://ftp.ietf.org/rfc/rfc2627.txt`, 1997.

[23] C.K. Wong, M. Gouda, and S. Lam. Secure Group Communications Using Key Graphs. In *Proceedings of the ACM SIGCOMM '98*, 1998.

Appendix

The proofs of the following lemmas are based on the same techniques used in [8]; the main tool is the following technical lemma.

Lemma 8 *Let k,n be integers with $1 \leq k \leq n$, and let K be a finite field. Consider a probability space with random variables $\vec{\alpha} \in K^{n\times 1}, \vec{\beta} = (\beta_1, \ldots, \beta_k)^T \in K^{k\times 1}, \vec{\gamma} \in K^{k\times 1}$, and $M \in K^{k\times n}$, such that $\vec{\alpha}$ is uniformly distributed over $K^n, \vec{\beta} = M\vec{\alpha} + \vec{\gamma}$, and for $1 \leq i \leq k$, the first i^{th} rows of M and $\vec{\gamma}$ are determined by $\beta_1, \ldots, \beta_{i-1}$. Then, conditioning on any fixed values of $\beta_1, \ldots, \beta_{k-1}$ such that the resulting matrix M has rank k, the value of β_k is uniformly distributed over K in the resulting conditional probability space.*

In what follows, we will denote with Coins the coin tosses of $\mathcal{A}$ and we define

$$\mathbf{X}_t \doteq X_{1,t} + wX_{2,t}, \quad \mathbf{Y}_t \doteq Y_{1,t} + wY_{2,t}, \quad \mathbf{Z}_t \doteq Z_{1,t} + wZ_{2,t}, \qquad t = 0 \ldots z.$$

Proof of the Lemma stated in Theorem 4

Lemma 9 $\Pr[T_4] = \Pr[T_3]$.

Proof: Consider the quantity $V := (\text{Coins}, w, \mathbf{Z}_1, \ldots, \mathbf{Z}_z, \sigma, r_1^*, r_2^*)$ and the value $\mathbf{Z}_0$. According to the specification of games $\mathbf{G}_2$ and $\mathbf{G}_3$, V and $\mathbf{Z}_0$ assume the same value in both games. Let us now consider the value $e^* = \log_{g_1} S^*$: unlike the previous two quantities, e^* assumes different values in the above two games. In particular, while in game $\mathbf{G}_2$ e^* contains information about the session key s_σ, in game $\mathbf{G}_3$ e^* is just a random value: let us denote with $[e^*]_2$ and $[e^*]_3$ the values of e^* in game $\mathbf{G}_2$ and game $\mathbf{G}_3$, respectively.

By definition of game $\mathbf{G}_2$, event T_2 solely depends on $(V, \mathbf{Z}_0, [e^*]_2)$; similarly, by definition of game $\mathbf{G}_3$, event T_3 solely depends on $(V, \mathbf{Z}_0, [e^*]_3)$. Moreover, event T_2 depends on $(V, \mathbf{Z}_0, [e^*]_2)$ according to the same functional dependence of event T_3 upon $(V, \mathbf{Z}_0, [e^*]_3)$. Therefore, to prove the lemma, it suffices to show that $(V, \mathbf{Z}_0, [e^*]_2)$ and $(V, \mathbf{Z}_0, [e^*]_3)$ have the same distribution.

According to the specification of game $\mathbf{G}_3$, $[e^*]_3$ is chosen uniformly over $\mathbb{Z}_q$, independently from V and $\mathbf{Z}_0$. Hence, to reach the thesis, it suffices to prove that the distribution of $[e^*]_2$, conditioned on V and $\mathbf{Z}_0$, is also uniform in $\mathbb{Z}_q$.

In game $\mathbf{G}_2$, the quantities $(V, \mathbf{Z}_0, [e^*]_2)$ are related according to the following matrix equation:

$$\begin{pmatrix} \mathbf{Z}_0 \\ [e^*]_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & w \\ r_1^* & wr_2^* \end{pmatrix}}_{M} \cdot \begin{pmatrix} Z_{1,0} \\ Z_{2,0} \end{pmatrix} + \begin{pmatrix} 0 \\ \log_{g_1} s_\sigma \end{pmatrix}$$

where $det(M) = w(r_2^* - r_1^*) \neq 0$, since $r_2^* \neq r_1^*$.
As soon as we fix the value of V, the matrix M is completely fixed, but the values $Z_{1,0}$ and $Z_{2,0}$ are still uniformly and independently distributed over $\mathbb{Z}_q$. Now, fixing a value for $\mathbf{Z}_0$ also fixes a value for s_σ; hence, by Lemma 8, we can conclude that the conditioned distribution of $[e^*]_2$, w.r.t. V and $\mathbf{Z}_0$, is also uniform over $\mathbb{Z}_q$. □

Proofs of Lemmas stated in Theorem 6

Lemma 10 $\Pr[T_4] = \Pr[T_3]$ *and* $\Pr[R_4] = \Pr[R_3]$.

Proof: Consider the quantity:

$$V := (\textsc{Coins}, \mathcal{H}, w, X_{1,0}, X_{2,0}, \ldots, X_{1,z}, X_{2,z}, Y_{1,0}, Y_{2,0}, \ldots, Y_{1,z}, Y_{2,z}, \mathbf{Z}_1, \ldots, \mathbf{Z}_z, \sigma, r_1^*, r_2^*)$$

and the value $\mathbf{Z}_0$. Introducing similar notations as in Lemma 9 and reasoning as above, we can notice that event T_3 solely depends on $(V, \mathbf{Z}_0, [e^*]_3)$ and that event T_4 solely depends on $(V, \mathbf{Z}_0, [e^*]_4)$. Moreover, event T_3 depends on $(V, \mathbf{Z}_0, [e^*]_3)$ according to the same functional dependence of event T_4 upon $(V, \mathbf{Z}_0, [e^*]_4)$. The same considerations hold for events R_3 and R_4. Therefore, to prove the lemma, it suffices to show that $(V, \mathbf{Z}_0, [e^*]_3)$ and $(V, \mathbf{Z}_0, [e^*]_4)$ have the same distribution.

According to the specification of game $\mathbf{G}_4$, $[e^*]_4$ is chosen uniformly over $\mathbb{Z}_q$, independently from V and $\mathbf{Z}_0$. Hence, to reach the thesis, it suffices to prove that the distribution of $[e^*]_3$, conditioned on V and $\mathbf{Z}_0$, is also uniform in $\mathbb{Z}_q$.

In game $\mathbf{G}_3$, the quantities $(V, \mathbf{Z}_0, [e^*]_3)$ are related according to the following matrix equation:

$$\begin{pmatrix} \mathbf{Z}_0 \\ [e^*]_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & w \\ r_1^* & wr_2^* \end{pmatrix}}_{M} \cdot \begin{pmatrix} Z_{1,0} \\ Z_{2,0} \end{pmatrix} + \begin{pmatrix} 0 \\ \log_{g_1} s_\sigma \end{pmatrix}$$

where $det(M) = w(r_2^* - r_1^*) \neq 0$, since $r_2^* \neq r_1^*$.
As soon as we fix the value of V, the matrix M is completely fixed, but the values $Z_{1,0}$ and $Z_{2,0}$ are still uniformly and independently distributed over $\mathbb{Z}_q$. Now, fixing a value for $\mathbf{Z}_0$ also fixes a value for s_σ; hence, by Lemma 8, we can conclude that the conditioned distribution of $[e^*]_3$, w.r.t. V and $\mathbf{Z}_0$, is also uniform over $\mathbb{Z}_q$. □

Lemma 11 *If $Q_{\mathcal{A}}(\lambda)$ is an upper bound on the number of decryption queries that A poses to the decryption algorithm, then* $\Pr[R_5] \leq \frac{Q_{\mathcal{A}}(\lambda)}{q}$.

Proof: In what follows, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, we will denote with $R_5^{(j)}$ the event that the j^{th} ciphertext $\langle i, \mathcal{T} \rangle$, submitted by $\mathcal{A}$ to the decryption oracle in game $\mathbf{G}_5$, fails to pass the test in step

$D4'$, but would have passed the test in step $D4$ in game $\mathbf{G}_2$. Besides, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, we will denote with $B_5^{(j)}$ the event that the j^{th} ciphertext is submitted to the decryption oracle before $\mathcal{A}$ received her challenge, and with $\hat{B}_5^{(j)}$ the event that the j^{th} ciphertext is submitted to the decryption oracle after $\mathcal{A}$ received her challenge. If we show that, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, $\Pr[R_5^{(j)} \mid B_5^{(j)}] \leq \frac{1}{q}$ and that $\Pr[R_5^{(j)} \mid \hat{B}_5^{(j)}] \leq \frac{1}{q}$, then the thesis will follow.

Claim.: $\Pr[R_5^{(j)} \mid B_5^{(j)}] \leq \frac{1}{q}$.
To prove this claim, fix $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$ and consider the quantities:

$$V := (\textsc{Coins}, \mathcal{H}, w, \mathbf{Z}_0, \ldots, \mathbf{Z}_z), \quad V' := (\mathbf{X}_0, \ldots, \mathbf{X}_z, \mathbf{Y}_0, \ldots, \mathbf{Y}_z).$$

These two quantities together contain all the randomness needed to determine the behavior of $\mathcal{A}$ and of all the oracles she interacts with, up to the moment that $\mathcal{A}$ performs the encryption query: once we fix V and V', we totally define how the adversary proceeds in her attack, before she receives her challenge back. Moreover, fixing V and V', the event $B_5^{(j)}$ is completely defined: given V and V', we say they are *relevant*, if the event $B_5^{(j)}$ occurs.

Hence, to reach the claim, it suffice to prove that the probability of event $R_5^{(j)}$, conditioned on any *relevant* values of V and V', is less then $1/q$.

Recall that the condition tested in step $D4'$ in game $\mathbf{G}_5$ is $(u_2 = u_1^w \wedge v_i = \bar{v}_i)$: since we are considering the case that the j^{th} query fails to pass the test in step $D4'$, but would have passed the test in step $D4$ of game $\mathbf{G}_2$, it must be the case that $v_i = \bar{v}_i$ but $u_2 \neq u_1^w$. Therefore, we only consider *relevant* values of V and V' such that $u_2 \neq u_1^w$. Taking the logs (base g_1), the condition $u_2 \neq u_1^w$ is equivalent to $r_1 \neq r_2$ and the condition $v_i = \bar{v}_i$ is equivalent to $\beta_i = \bar{\beta}_i$, where $\bar{\beta}_i \doteq \log_{g_1} \bar{v}_i = r_1 X_{1,i} + w r_2 X_{2,i} + \alpha r_1 Y_{1,i} + \alpha w r_2 Y_{2,i}$ and $\beta_i \doteq \log_{g_1} v_i = \mathsf{LI}(0, \ldots, z; \log_{g_1} v_0, \ldots, \log_{g_1} v_z)(i)$. Notice that $\bar{\beta}_i$ can be expressed in terms of the vector $(X_{1,0}, X_{2,0}, \ldots, X_{1,z}, X_{2,z}, Y_{1,0}, Y_{2,0}, \ldots, Y_{1,z}, Y_{2,z})^T$; indeed, $X_{1,i} = \mathsf{LI}(0, \ldots, z; X_{1,0}, \ldots, X_{1,z})(i) = \sum_{t=0}^{z} (X_{1,t} \cdot \lambda_t(i))$, and similar relations hold for $X_{2,i}, Y_{1,i}$ and $Y_{2,i}$. Therefore, by means of some matrix manipulation, we can write:

$$\bar{\beta}_i = \vec{\delta} \cdot (X_{1,0}, X_{2,0}, \ldots, X_{1,z}, X_{2,z}, Y_{1,0}, Y_{2,0}, \ldots, Y_{1,z}, Y_{2,z})^T$$

where $\vec{\delta} \equiv (\delta_0, \delta_1, \ldots, \delta_{2z}, \delta_{2z+1}, \delta_{2z+2}, \delta_{2z+3}, \ldots, \delta_{4z+2}, \delta_{4z+3})$ is defined as:

$$\vec{\delta} \doteq (r_1\lambda_0(i), w r_2 \lambda_0(i), \ldots, r_1 \lambda_z(i), w r_2 \lambda_z(i), \alpha r_1 \lambda_0(i), \alpha w r_2 \lambda_0(i), \ldots, \alpha r_1 \lambda_z(i), \alpha w r_2 \lambda_z(i)).$$

In game $\mathbf{G}_5$, the random values defined above are related according to the following matrix equation:

$$\begin{pmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_z \\ \mathbf{Y}_0 \\ \vdots \\ \mathbf{Y}_z \\ \bar{\beta}_i \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & w & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & \vdots & & & & \vdots \\ 0 & 0 & \ldots & 1 & w & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & w & \ldots & 0 & 0 \\ \vdots & & & & \vdots & \vdots & & & & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 & w \\ \delta_0 & \delta_1 & \ldots & \delta_{2z} & \delta_{2z+1} & \delta_{2z+2} & \delta_{2z+3} & \ldots & \delta_{4z+2} & \delta_{4z+3} \end{pmatrix}}_{M} \cdot \begin{pmatrix} X_{1,0} \\ X_{2,0} \\ \vdots \\ X_{1,z} \\ X_{2,z} \\ Y_{1,0} \\ Y_{2,0} \\ \vdots \\ Y_{1,z} \\ Y_{2,z} \end{pmatrix}$$

We want to show that the rank of the matrix M is $2z+3$. Clearly, the first $2z+2$ rows are linearly independent; to see why the last row (i.e. the vector $\vec{\delta}$) is independent from the others, notice that the only way to obtain δ_0 is by multiplying the first row by $r_1\lambda_0(i)$: doing so, the second component of δ results to be $wr_1\lambda_0(i)$. But since $\delta_1 = wr_2\lambda_0(i)$, this implies that $r_1 = r_2$, contradicting the assumption that the query fails to pass the test in step $D4'$ in game $\mathbf{G}_5$.

As soon as we fix V, the first $2z+2$ rows of matrix M are fixed, but the values $X_{1,0}, X_{2,0}, \ldots, Y_{1,z}$, $Y_{2,z}$ are still uniformly and independently distributed over $\mathbb{Z}_q$; as for $\vec{\delta}$, its value is still undetermined, since r_1, r_2 and i are not yet fixed. Now, fixing a value for V' such that V and V' are *relevant* and that $r_1 \neq r_2$, determines the value of the j^{th} query (and hence the value of $\vec{\delta}$), along with the values $\mathbf{X}_0, \ldots, \mathbf{X}_z, \mathbf{Y}_0, \ldots, \mathbf{Y}_z$ and $\bar{\beta}_i$. Therefore, by Lemma 8, we can conclude that the distribution of $\bar{\beta}_i$, conditioned on relevant values of V and V', is uniform over $\mathbb{Z}_q$; since conditioning on any fixed, relevant value of V and V', β_i is just a single point in $\mathbb{Z}_q$, it follows that $\Pr[\beta_i = \bar{\beta}_i] = \frac{1}{q}$.

Claim.: $\Pr[R_5^{(j)} \mid \hat{B}_5^{(j)}] \leq \frac{1}{q}$.
To prove this claim, fix $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$ and consider the quantities:

$$V := (\text{Coins}, \mathcal{H}, w, \mathbf{Z}_0, \ldots, \mathbf{Z}_z, r_1^*, r_2^*, e^*), \quad V' := (\mathbf{X}_0, \ldots, \mathbf{X}_z, \mathbf{Y}_0, \ldots, \mathbf{Y}_z, \beta_i^*)$$

where $\beta_i^* \doteq \log_{g_1} v_i^* = \mathsf{LI}(0, \ldots, z; \log_{g_1} v_0^*, \ldots, \log_{g_1} v_z^*)(i)$ and $i > z$. Notice that by the specification of the encryption oracle used in game $\mathbf{G}_5$, it holds that: $\log_{g_1} v_t^* = r_1^* X_{1,t} + wr_2^* X_{2,t} + \alpha^* r_1^* Y_{1,t} + \alpha^* wr_2^* Y_{2,t}$, $t = 0, \ldots, z$. Therefore, we can write:

$$\beta_i^* = \sum_{t=0}^{z} \lambda_t(i)(r_1^* X_{1,t} + wr_2^* X_{2,t} + \alpha^* r_1^* Y_{1,t} + \alpha^* wr_2^* Y_{2,t}).$$

Together, V and V' contain all the parameters needed to determine the behavior of $\mathcal{A}$ and of all the oracles she interacts with: once we fix V and V', we totally define how the adversary proceeds in the *entire* attack. Moreover, fixing V and V', the event $\hat{B}_5^{(j)}$ is completely defined: given V and V', we say they are *relevant*, if the event $\hat{B}_5^{(j)}$ occurs.

Hence, to reach the claim, it suffices to prove that the probability of event $R_5^{(j)}$, conditioned on any *relevant* values of V and V', is less then $1/q$.

As shown above, we can consider just relevant values of V and V' for which it holds that $u_2 \neq u_1^w$. Reasoning as in the previous case, and maintaining the notation introduced there, the random values defined above are related according to the following matrix equation:

$$\begin{pmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_z \\ \mathbf{Y}_0 \\ \vdots \\ \mathbf{Y}_z \\ \beta_i^* \\ \bar{\beta}_i \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & w & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & \vdots & & & & \vdots \\ 0 & 0 & \ldots & 1 & w & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & w & \ldots & 0 & 0 \\ \vdots & & & & \vdots & \vdots & & & & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 & w \\ \delta_0^* & \delta_1^* & \ldots & \delta_{2z}^* & \delta_{2z+1}^* & \delta_{2z+2}^* & \delta_{2z+3}^* & \ldots & \delta_{4z+2}^* & \delta_{4z+3}^* \\ \delta_0 & \delta_1 & \ldots & \delta_{2z} & \delta_{2z+1} & \delta_{2z+2} & \delta_{2z+3} & \ldots & \delta_{4z+2} & \delta_{4z+3} \end{pmatrix}}_{M} \cdot \begin{pmatrix} X_{1,0} \\ X_{2,0} \\ \vdots \\ X_{1,z} \\ X_{2,z} \\ Y_{1,0} \\ Y_{2,0} \\ \vdots \\ Y_{1,z} \\ Y_{2,z} \end{pmatrix}$$

where $\vec{\delta}^* \equiv (\delta_0^*, \delta_1^*, \ldots, \delta_{2z}^*, \delta_{2z+1}^*, \delta_{2z+2}^*, \delta_{2z+3}^*, \ldots, \delta_{4z+2}^*, \delta_{4z+3}^*)$ is defined as:

$$\vec{\delta}^* \doteq (r_1^*\lambda_0(i), wr_2^*\lambda_0(i), \ldots, r_1^*\lambda_z(i), wr_2^*\lambda_z(i), \alpha^* r_1^*\lambda_0(i), \alpha^* wr_2^*\lambda_0(i), \ldots, \alpha^* r_1^*\lambda_z(i), \alpha^* wr_2^*\lambda_z(i)).$$

We want to show that the rank of the matrix M is $2z+4$. Clearly, the first $2z+2$ rows of M are all linear independent. Moreover, as shown in the previous claim, both β_i^* and $\bar{\beta}_i$ are linearly independent from the first $2z+2$ rows of M. Firstly, notice that the assumption that the j^{th} query $\langle i, \mathcal{T}\rangle$ is rejected in step $D4'$ of game $\mathbf{G}_5$, implies not only $\neg\Re_i(\mathcal{T}, \mathcal{T}^*)$, but also that $\mathcal{T}$ passed the *special rejection rule*; furthermore, we may assume that $\alpha \neq \alpha^*$, since otherwise the only way $\mathcal{T}$ may have passed the *special rejection rule* is that $\langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z})\rangle = \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*})\rangle$. But this on one hand entails $\vec{\delta}^* = \vec{\delta}$, i.e. $\beta_i^* = \bar{\beta}_i$, whereas on the other hand implies that $\beta_i \neq \beta_i^*$ (because otherwise $\mathcal{T}$ and $\mathcal{T}^*$ would be $\Re_i$-related). Thus, if $\alpha = \alpha^*$ then $\beta_i \neq \bar{\beta}_i$, contradicting the assumption that the j^{th} query $\langle i, \mathcal{T}\rangle$ would have passed the test in step $D4$ in game $\mathbf{G}_2$.

In order to show that $\vec{\delta}$ is linearly independent from the first $2z+3$ rows, observe that the only way to obtain δ_0 is by multiplying the first row by $(r_1 - r_1^*)\lambda_0(i)$ and $\vec{\delta}^*$ by 1; similarly, to obtain δ_{2z+2} as a linear combination of the other elements in its column, we need to multiply the $(z+2)^{\text{th}}$ row by $\alpha(r_1 - r_1^*)\lambda_0(i)$ and $\vec{\delta}^*$ by $\frac{\alpha}{\alpha^*}$: since $\alpha \neq \alpha^*$, $\frac{\alpha}{\alpha^*} \neq 1$ and so, $\vec{\delta}$ is linearly independent from all the other rows.

As soon as we fix V, the first $2z+2$ rows of matrix M are fixed, but the values $X_{1,0}, X_{2,0}, \ldots, Y_{1,z}$, $Y_{2,z}$ are still uniformly and independently distributed over $\mathbb{Z}_q$; as for δ^* and δ, their values are still undetermined, since r_1^*, r_2^*, r_1, r_2 and i, are not yet fixed. Now, fixing a value for V' such that V and V' are *relevant* and that $r_1 \neq r_2$, also fixes the last 2 rows of matrix M along with the values $\mathbf{X}_0, \ldots, \mathbf{X}_z, \mathbf{Y}_0, \ldots, \mathbf{Y}_z$ and β_i^*; hence, by Lemma 8, we can conclude that the distribution of $\bar{\beta}_i$, conditioned on relevant values of V and V', is also uniform over $\mathbb{Z}_q$; since conditioning on any fixed, relevant values of V and V', β_i is just a single point in $\mathbb{Z}_q$, it follows that $\Pr[\beta_i = \bar{\beta}_i] = \frac{1}{q}$. □

Proofs of Lemmas stated in Theorem 7

Lemma 12 $\Pr[T_4] = \Pr[T_3]$ *and* $\Pr[R_4] = \Pr[R_3]$.

Proof: Consider the quantities:

$$V := (\textsc{Coins}, \mathcal{H}, w, X_{1,0}, X_{2,0}, \ldots, X_{1,z}, X_{2,z}, Y_{1,0}, Y_{2,0}, \ldots, Y_{1,z}, Y_{2,z}, \mathbf{Z}_1, \ldots, \mathbf{Z}_z, \sigma, r_1^*, r_2^*, k).$$

and the value $\mathbf{Z}_0$. We can repeat the same considerations stated in Lemma 10: the only difference is that the quantities $(V, \mathbf{Z}_0, [e^*]_3)$ characterizing game $\mathbf{G}_3$ are related according to the following slightly different matrix equation:

$$\begin{pmatrix} \mathbf{Z}_0 \\ [e^*]_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & w \\ r_1^* & wr_2^* \end{pmatrix}}_{M} \cdot \begin{pmatrix} Z_{1,0} \\ Z_{2,0} \end{pmatrix} + \begin{pmatrix} 0 \\ \log_{g_1}(s_\sigma \| k) \end{pmatrix}$$

For the same reasons seen in Lemma 10, as soon as we fix a value for V, the matrix M is completely fixed, as well as the value of k, but the values $Z_{1,0}$ and $Z_{2,0}$ are still uniformly and independently distributed over $\mathbb{Z}_q$. Now, fixing a value for $\mathbf{Z}_0$ also fixes a value for s_σ and hence for $\log_{g_1}(s_\sigma \| k)$; thus, by Lemma 8, the conditioned distribution of $[e^*]_3$, w.r.t. V and $\mathbf{Z}_0$, is also uniform over $\mathbb{Z}_q$. □

Lemma 13 *If $Q_{\mathcal{A}}(\lambda)$ is an upper bound on the number of decryption queries that A poses to the decryption algorithm, then* $\Pr[R_6] \leq \frac{Q_{\mathcal{A}}(\lambda)}{q}$.

Proof: In what follows, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, we will denote with $R_6^{(j)}$ the event that the j^{th} ciphertext $\langle i, \mathcal{T} \rangle$, submitted by $\mathcal{A}$ to the decryption oracle in game $\mathbf{G}_6$, fails to pass the test in step $D4'$, but would have passed both tests in step $D4$ and in step $D8$ in game $\mathbf{G}_2$. Besides, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, we will denote with $B_6^{(j)}$ the event that the j^{th} ciphertext is submitted to the decryption oracle before $\mathcal{A}$ received her challenge, and with $\hat{B}_6^{(j)}$ the event that the j^{th} ciphertext is submitted to the decryption oracle after $\mathcal{A}$ received her challenge. If we show that, for $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$, $\Pr[R_6^{(j)} \mid B_6^{(j)}] \leq \frac{1}{q}$ and that $\Pr[R_6^{(j)} \mid \hat{B}_6^{(j)}] \leq \frac{1}{q}$, then the thesis will follow.

Claim.: $\Pr[R_6^{(j)} \mid B_6^{(j)}] \leq \frac{1}{q}$.
This proof closely follows the one presented in Lemma 11, so we omit the details here.

Claim.: $\Pr[R_6^{(j)} \mid \hat{B}_6^{(j)}] \leq \frac{1}{q}$.
To prove this claim we proceed like in Lemma 11, fixing $1 \leq j \leq Q_{\mathcal{A}}(\lambda)$ and considering the quantities:

$$V := (\text{Coins}, \mathcal{H}, w, \mathbf{Z}_0, \ldots, \mathbf{Z}_z, r_1^*, r_2^*, e^*), \quad V' := (\mathbf{X}_0, \ldots, \mathbf{X}_z, \mathbf{Y}_0, \ldots, \mathbf{Y}_z, \beta_i^*, k)$$

where we are maintaining all the notations introduced above.

Again, we can repeat exactly the same construction utilized in Lemma 11: the only difference from the argument presented there is in the considerations aiming at showing that we can assume that $\alpha \neq \alpha^*$; thus, we only need to justify this assumption in the new scenario, and the claim will follow.

Under the assumptions that the j^{th} query $\langle i, \mathcal{T} \rangle$ is rejected in step $D4'$ of game $\mathbf{G}_6$ but would have been decrypted as valid in game $\mathbf{G}_2$, we can deduce that $\mathcal{T}$ passed both the *second special rejection rule* and the *special rejection rule*. We may also assume that $\alpha \neq \alpha^*$, since otherwise the only way that $\mathcal{T}$ may have passed the *special rejection rule* is that $\langle S, u_1, u_2, (j_1, H_{j_1}), \ldots, (j_z, H_{j_z}) \rangle = \langle S^*, u_1^*, u_2^*, (j_1^*, H_{j_1^*}), \ldots, (j_z^*, H_{j_z^*}) \rangle$; but since $\mathcal{T}$ must differ from the challenge $\mathcal{T}^*$, then it must be the case that $(v_0, \ldots, v_z) \neq (v_0^*, \ldots, v_z^*)$, and so, from the fact that $\mathcal{T}$ passed the *second special rejection rule* we get that $\tau \neq \mathsf{MAC}_{k^*}(v_0, \ldots, v_z)$, thus contradicting the assumption that the j^{th} query would have been decrypted in game $\mathbf{G}_2$ (since the test in step $D8$, for the validity of the tag τ, would have failed). □