Remarks about the Security Proofs of Boneh-Franklin's Identity-Based Encryption

Yu Chen
Institute of Software, School of Electronics Engineering and Computer Science
Peking University, Beijing, China
cycosmic@gmail.com

Abstract

Boneh and Franklin constructed the first practical Identity-Based Encryption (IBE) scheme in 2001. They also defined a formal security model for IBE and proved their scheme (BF-IBE) to be secure in the random oracle model under the computational Bilinear Diffie-Hellman assumption. However, few years later, Galindo [10] pointed out a flawed step in its proof of chosen ciphertext attack (CCA) security and claimed that it is possible to fixed it without changing the original scheme and underlying assumption. In the same paper, Galindo provided a revised proof with a looser security reduction. Shortly afterwards, Nishioka [17] attempted to extend Galindo's idea to achieve a tighter security reduction. Unfortunately, we find that there are some lapses in their mending works, which make their proofs not infallible. In this paper, except pointing out the defects in the aforementioned proofs, we also illustrate how to fix their proofs. Besides, we provide a security proof for BF-IBE in selective identity model. Furthermore, we prove BF-IBE is fully secure in the random oracle model if confining its identity space to a finite set.

1 Introduction

Identity-Based Encryption (IBE) allows a party to encrypt a message using the recipient's identity as a public key. Such property simplifies key management and avoids the use of digital certificates. This can be very useful in applications such as email system where the recipient is often off-line and unable to present a public-key certificate while the sender encrypts a message.

Since Shamir proposed the concept of IBE in 1984 [19], various Identity-Based signature (IBS) and authentication (IBA) schemes have been proposed, but secure and fully functional IBE scheme was not found until Boneh and Franklin [3], Cocks [8] and Sakai *et al.* [18] presented three IBE schemes in 2001, respectively. Among those solutions, Boneh and Franklin's scheme (BF-IBE) happen to be the most practical one. In order to prove the security of BF-IBE, Boneh and Franklin [3] introduced new security model to fit the Identity-Based setting, then proved its security assuming the hardness of *computational* Bilinear Diffie-Hellman problem. Because BF-IBE is the first fully functional practical IBE scheme, even though its security is given in the heuristic model (the random oracle model [16]), it still has had a great influence on later designs and analyses of cryptographic schemes. Numerous schemes [1] [5] [13] [14] [15] are based on BF-IBE schemes.

The original security proof of BF-IBE was long believed correct until 2005, Galindo [10] pointed out a flawed step in one security reduction for CCA security. Galindo claimed that the flawed step could be fixed by his new security reduction without changing both the scheme and the underlying assumption if the efficiency of the security reduction is sacrificed. In the same year, Nishioka [17] enhanced Galindo's idea to provide another proof with tighter security reduction. Up to present, no one doubts the correctness of their mending proofs.

1.1 Our contributions

We first re-examine the flawed steps in the original proof of BF-IBE exhibited in [4] and analyze the reason why it fails. Then we identify some questionable issues in the subsequent revised security proofs proposed by Galindo [10] and Nishioka [17], respectively. Both of their proofs begin with a doubtable hypothesis that the simulator know the exact number of queries q_H that the adversary will make even at the beginning of the CCA game. Apart from this common issue, for Galindo's proof the simulation algorithm is not well defined, which leads the probability of perfect simulation is immeasurable. In addition, in the challenge step the behavior of the adversary is inconsistency to the definition of CCA security. For Nishioka's proof, the computation of the probability that the simulator does not aborts

is not right. We show that their proofs could be fixed by a minor modification, i.e. replacing the particular of q_H with the upper bound of q_H. As the last contribution, we provide a security proof of BF-IBE in selective identity model. Moreover, we show that BF-IBE can be proved to be fully secure in the random oracle model if confining its identity space to a finite set.

1.2 Organization

In next section 2, we give the background information about the related security definitions and assumptions. In Section 3 we briefly review the BF-IBE scheme and its original security proof. In Section 4 we identify some questionable issues in the subsequent revised proofs proposed by Galindo [10] and Nishioka [17] and illustrate how to fix them. In Section 5 we prove BF-IBE is secure in the selective-ID model. Finally, we conclude the paper in Section 6.

2 Preliminaries

We briefly review the groups with efficiently computable bilinear maps that will be used throughout the paper. For more details, we recommend the reader to previous literature [4].

Bilinear Map. $\mathbb{G}_1$ and $\mathbb{G}_2$ be two groups of large prime order q. A map $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is said as an admissible bilinear map if the following three properties hold.

1. Bilinear. $e(aP, bQ) = e(P, Q)^{ab}$ for all $P, Q \in \mathbb{G}_1$ and all $a, b \in \mathbb{Z}_q^*$.
2. Non-degenerate. $e(P, P) \neq 1$.
3. Computable. There is an efficient algorithm to compute $e(P, Q)$ for any $P, Q \in \mathbb{G}_1$.

Bilinear Diffie-Hellman (BDH) Parameter Generator. A BDH parameter generator $\mathcal{G}$ is an algorithm which takes a security parameter $k \in \mathbb{Z}^+$ as input and outputs two groups of prime order q and an admissible bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$. We describe it as $\mathcal{G}(1^k) = \langle q, \mathbb{G}_1, \mathbb{G}_2, e\rangle$.

Computational Bilinear Diffie-Hellman Problem. Given the tuple (P, aP, bP, cP) for some $a, b, c \in \mathbb{Z}_q^*$, $P \in \mathbb{G}_1$, to compute the $e(P, P)^{abc} \in \mathbb{G}_2$. An adversary $\mathcal{A}$ is said to have at least advantage ϵ in solving CBDH problem if $\Pr\left[\mathcal{A}(P, aP, bP, cP) = e(P, P)^{abc}\right] \geq \epsilon$.

2.1 Security Notions

Recall that an IBE system consists of four algorithms [19] [4]: Setup, Extract, Encrypt, and Decrypt. The Setup algorithm generates system parameters params and a master secret master-key. The Extract algorithm uses the master-key to generate the private key corresponding to a given identity. The Encrypt algorithm encrypts messages for a given identity (using the system parameters) and the Decrypt algorithm decrypts ciphertext using the private key. The message space is $\mathcal{M}$. The ciphertext space is $\mathcal{C}$.

Chosen Ciphertext Security for Identity-Based Encryption. An IBE scheme $\mathcal{E}$ is said to be secure against adaptively chosen ciphertext attack (IND-ID-CCA) if no probabilistic polynomial time (PPT) algorithm $\mathcal{A}$ has a non-negligible advantage against the challenger in the following game:

Setup. The challenger runs the Setup algorithm. It gives the adversary the resulting system parameters params and keeps the master-key to itself.

Phase 1. The adversary $\mathcal{A}$ issues queries $q_1, \ldots, q_m$ where query q_i is one of:

- Extraction query $\langle \mathsf{ID}_i\rangle$. The challenger responds by running algorithm Extract to generate the private key d_i corresponding to ID_i and sends d_i to $\mathcal{A}$.
- Decryption query $\langle \mathsf{ID}_i, C_i\rangle$. The challenger responds by running algorithm Extract to generate the private key d_i corresponding to ID_i. It then runs algorithm Decrypt to decrypt the ciphertext C_i using the private key d_i and sends the resulting plaintext to $\mathcal{A}$.

These queries may be asked adaptively, that is, each query q_i may depends on the replies to $q_1, \ldots, q_{i-1}$.

Challenge. Once the adversary $\mathcal{A}$ decides that Phase 1 is over it outputs two equal length plaintexts $M_0, M_1 \in \mathcal{M}$ and an identity ID_{ch} on which it wishes to be challenged. The only constraint is that ID_{ch} did not appear in any private key extraction query in Phase 1. The challenger picks a random bit $c \in \{0, 1\}$ and sets $C = \mathsf{Encrypt}(\mathsf{params}, \mathsf{ID}, M_c)$. It sends C as the challenge to $\mathcal{A}$.

Phase 2. $\mathcal{A}$ issues more queries $q_{m+1}, \ldots, q_r$ where q_i is one of:

- Extraction query $\langle \mathsf{ID}_i\rangle \neq \langle \mathsf{ID}\rangle$. Challenger responds as in Phase 1.
- Decryption query $\langle \mathsf{ID}_i, C_i\rangle \neq \langle \mathsf{ID}, C\rangle$. Challenger responds as in Phase 1.

These queries may be asked adaptively as in Phase 1.

Guess. Finally, $\mathcal{A}$ outputs a guess $c' \in \{0, 1\}$ and wins the game if $c = c'$.

We refer to such an adversary $\mathcal{A}$ as an IND-ID-CCA adversary, and define the adversary $\mathcal{A}$'s advantage over the scheme $\mathcal{E}$ by $Adv_{\mathcal{E},\mathcal{A}}(k) = \left|\Pr[c = c'] - \frac{1}{2}\right|$, where k is the security parameter. The probability is over the random bits used by the challenger and the adversary.

Definition 2.1. *An IBE scheme $\mathcal{E}$ is* IND-ID-CCA *secure*

if for any PPT IND-ID-CCA *adversary* $\mathcal{A}$ *the advantage* $Adv_{\mathcal{E},\mathcal{A}}(k)$ *is negligible.*

Selective-ID model. Boneh and Franklin [3] defined the adaptive chosen ciphertext security for IBE systems by the above game. We refer to it as full IBE security model. In this model, the adversary can issue both adaptive chosen ciphertext queries and adaptive chosen identity queries. Eventually, the adversary adaptively chooses the identity it wishes to attack and asks for a semantic security challenge for this identity. Canetti, Halevi, and Katz [6] [7] defined a slightly weaker security model, called selective-ID security model, in which the adversary must commit ahead of time (non-adaptively) to the identity it intends to attack. More precisely, it is defined using the following game:

Init. The adversary outputs an identity ID_{ch} where it wishes to be challenged.

Setup and **Phase 1** are same as in IND-ID-CCA game.

Phase 1. Same as in IND-ID-CCA game.

Challenge. Once the adversary decides that Phase 1 is over it outputs two equal length plaintexts M_0, M_1 on which it wishes to be challenged. The challenger picks a random bit $c \in \{0,1\}$ and sets the challenge ciphertext to $C = \mathsf{Encrypt}(\mathsf{params}, \mathsf{ID}_{ch}, M_c)$. It sends C as the challenge to the adversary.

Phase 2 and **Guess** are same as in IND-ID-CCA game.

We refer to such an adversary $\mathcal{A}$ as an IND-sID-CCA adversary. We define the advantage of the adversary $\mathcal{A}$ over scheme $\mathcal{E}$ by $Adv_{\mathcal{E},\mathcal{A}}(k) = \left|\Pr[c = c'] - \frac{1}{2}\right|$. The probability is over the random bit used by the challenger and the adversary.

Definition 2.2. *An IBE system* $\mathcal{E}$ *is* IND-sID-CCA *secure if for any PPT* IND-sID-CCA *adversary* $\mathcal{A}$ *the advantage* $Adv_{\mathcal{E},\mathcal{A}}(k)$ *is negligible.*

3 Boneh-Franklin's IBE Scheme

In this section, we first briefly review the BF-IBE scheme, and then investigate its original proof [3]. Boneh and Franklin named their full scheme as FullIdent. In order to make the presentation easier, they introduced the BasicIdent and two public key encryption (PKE) scheme called BasicPub and BasicPub^{hy}. BasicIdent which has only CPA security, is a simplified version of FullIdent, BasicPub is a PKE scheme derived from BasicIdent, and BasicPub^{hy} is a PKE scheme obtained by applying the Fujisaki-Okamoto conversion [11] to BasicPub. Here follows the description of FullIdent.

Setup. Given a security number $k \in \mathbb{Z}^+$, the algorithm run $\mathcal{G}$ generate $\langle q, \mathbb{G}_1, \mathbb{G}_2, e\rangle$. Choose a random generator $P \in \mathbb{G}_1$. Pick a random $s \in \mathbb{Z}_q^*$ as the master secret and set $P_{pub} = sP$. Choose four cryptographic hash function $H_1 : \{0,1\}^* \to \mathbb{G}_1$, $H_2 : \mathbb{G}_2 \to \{0,1\}^n$ for some n, $H_3 : \{0,1\}^n \times \{0,1\}^n \to \mathbb{Z}_q^*$, and $H_4 : \{0,1\}^n \to \{0,1\}^n$. The system parameters are $\mathsf{params} = \langle q, \mathbb{G}_1, \mathbb{G}_2, e, n, P, P_{pub}, H_i\rangle$. The master-key is $s \in \mathbb{Z}_q^*$.

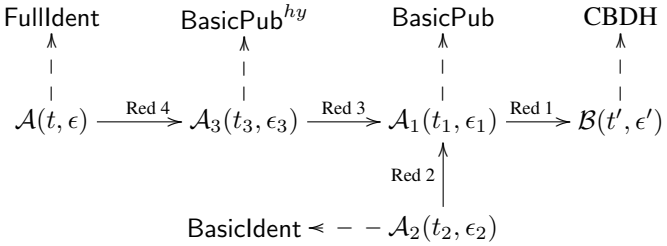
Extract. For a given string $\mathsf{ID} \in \{0,1\}^*$ the algorithm computes $Q_{\mathsf{ID}} = H_1(\mathsf{ID}) \in \mathbb{G}_1^*$, and sets the private key d_{ID} to be $d_{\mathsf{ID}} = sQ_{\mathsf{ID}}$ where s is the master-key.

Encrypt. To encrypt $M \in \mathcal{M}$ under the public key ID do the following: (1) compute $Q_{\mathsf{ID}} = H_1(\mathsf{ID}) \in \mathbb{G}_1^*$, (2) choose a random $\sigma \in \{0,1\}^n$, (3) set $r = H_3(\sigma, M)$, and set the ciphertext to be $C = \langle rP, \sigma \oplus H_2(g_{\mathsf{ID}}^r), M \oplus H_4(\sigma)\rangle$ where $g_{\mathsf{ID}} = e(Q_{\mathsf{ID}}, P_{pub}) \in \mathbb{G}_2$.

Decrypt. Let $C = \langle U, V, W\rangle$ be a ciphertext encrypted using the public key ID. If $U \notin \mathbb{G}_1^*$ reject the ciphertext. To decrypt C using the private key $d_{\mathsf{ID}} \in \mathbb{G}_1^*$ do: (1) Compute $V \oplus H_2(e(d_{\mathsf{ID}}, U)) = \sigma$. (2) Compute $W \oplus H_4(\sigma) = M$. (3) Set $r = H_3(\sigma, M)$. Test that $U = rP$. If not, reject the ciphertext. (4) Output M as the decryption of C.

This completes the description of FullIdent.

The security reductions from CBDH assumption to FullIdent and BasicIdent follows the diagram below.



The following results are presented in [4]. Hereafter, q_E, q_D, and q_{H_i} denote the number of extraction, decryption and random oracle H_i queries, respectively.

Reduction 1. Suppose there is an IND-CPA adversary $\mathcal{A}_1$ has the advantage $\epsilon(k)$ against BasicPub and $\mathcal{A}_1$ makes at most q_{H_2} queries to the random oracle H_2. Then there is an algorithm $\mathcal{B}$ that solves the CBDH problem with advantage at least $2\epsilon(k)/q_{H_2}$ in running time $O(time(\mathcal{A}_1))$.

Reduction 2. Suppose there is an IND-ID-CPA adversary $\mathcal{A}_2$ that has advantage $\epsilon(k)$ against BasicIdent and makes at most q_E private key extraction queries, and at most q_{H_2} queries to the random oracle H_2. Then there is an IND-CPA adversary $\mathcal{A}_1$ against BasicPub with advantage at least $\epsilon(k)/e(1+q_E)$ in running time $O(time(\mathcal{A}_2))$. Here $e \approx 2.71$ is the base of the natural logarithm.

From Reduction 1 and Reduction 2, we get:

Result 1. BasicIdent is IND-ID-CPA secure assuming the CBDH is hard in groups generated by $\mathcal{G}$. Concretely, suppose there is an IND-ID-CPA adversary $\mathcal{A}_2$ that has advantage $\epsilon(k)$ against BasicIdent. If $\mathcal{A}_2$ makes at most $q_E > 0$ private key extraction queries and q_{H_2} hash queries to H_2.

Then there is an algorithm $\mathcal{B}$ that solves CBDH with advantage at least $\frac{2\epsilon(k)}{e(1+q_E)\cdot q_{H_2}}$.

Reduction 3. Using the Fujisaki-Okamoto transformation Boneh and Franklin introduce $\mathsf{BasicPub}^{hy}$ which is IND-CCA secure. Suppose there is an IND-CCA adversary $\mathcal{A}_3$ that has advantage $\epsilon(k)$ against $\mathsf{BasicPub}^{hy}$ and makes at most q_D decryption queries, and at most q_{H_3}, q_{H_4} queries to the random oracles H_3, H_4 respectively. Then there exists an IND-CPA adversary $\mathcal{A}_1$ against BasicPub with advantage at least $[(\epsilon(k)+1)(1-2/q)^{q_D}-1]/2(q_{H_3}+q_{H_4})$ in running time $O(time(\mathcal{A}_3))$.

Reduction 4. Suppose there is an IND-ID-CCA adversary $\mathcal{A}$ that has advantage $\epsilon(k)$ against FullIdent. Suppose $\mathcal{A}$ makes at most q_E private key extraction queries, at most q_D decryption queries, and at most q_{H_1} queries to the random oracle H_1. Then there exists an IND-CCA adversary $\mathcal{A}_3$ against $\mathsf{BasicPub}^{hy}$ with advantage at least $\epsilon(k)/e(1+q_E+q_D)$ in running time $O(time(\mathcal{A}))$.

From Reduction 1, Reduction 3 and Reduction 4, we have:

Result 2. FullIdent is IND-ID-CCA secure assuming CBDH is hard in groups generated by $\mathcal{G}$. Concretely, suppose there is an IND-ID-CPA adversary $\mathcal{A}$ that has advantage $\epsilon(k)$ against BasicIdent. If $\mathcal{A}$ makes at most $q_E > 0$ private key extraction queries, at most q_D decryption queries, and at most q_{H_2}, q_{H_3}, q_{H_4} hash queries to H_2, H_3, H_4, respectively. Then there is an algorithm $\mathcal{B}$ that solves CBDH with advantage at least $\left[\frac{\epsilon(k)}{e(1+q_E+q_D)+1}(1-2/q)^{q_D}-1\right]/q_{H_2}(q_{H_3}+q_{H_4})$.

3.1 Analysis of Reduction 4 in BF-IBE

The aim of that Reduction 4 is constructing an IND-CCA adversary $\mathcal{A}_3$ against $\mathsf{BasicPub}^{hy}$ by interacting with an IND-ID-CCA adversary $\mathcal{A}$ against FullIdent. Next we list two lapses in Reduction 4 of BF-IBE, which is the Lemma 4.6 [4].

- **Issue 1**. In Phase 1, when $\mathcal{A}$ issues a decryption query $\langle \mathsf{ID}_i, C_i\rangle$, where $C_i = \langle U_i, V_i, W_i\rangle = \langle rP, \sigma \oplus H_2(e(Q_i, P_{pub})^r), M \oplus H_4(\sigma)\rangle$. According to the above algorithm, if $coin_i = 1$, $\mathcal{A}_3$ will modify C_i as $C_i' = \langle U_i', V_i', W_i'\rangle = \langle b_iU_i, V_i, W_i\rangle$ and then relay C_i' to its challenger. When the challenger decrypts C_i' using the private key d_{ID}, it does:
  1. Compute $V_i' \oplus H_2(e(d_{\mathsf{ID}}, U_i')) = V_i \oplus H_2(e(d_{\mathsf{ID}}, b_iU_i)) = \sigma \oplus H_2(e(Q_i, P_{pub})^r) \oplus H_2(e(sQ_{\mathsf{ID}}, b_irP)) = \sigma$. This step recovers the random chosen $\sigma \in \{0,1\}^n$ exactly.
  2. Compute $W_i' \oplus H_4(\sigma) = W_i \oplus H_4(\sigma) = M \oplus H_4(\sigma) \oplus H_4(\sigma) = M$. This step recovers the original plaintext M exactly.
  3. Set $r = H_3(\sigma, M)$. Test whether $U_i' = rP$. Note that b_i is randomly chosen from $\mathbb{Z}_q^*$ and H_3 is a random oracle model not controlled by $\mathcal{A}_3$. These facts imply that the probability of $H_3(\sigma, M) \neq b_ir$ is $1 - 1/q$, and therefore challenger will reject the modified ciphertext with overwhelming probability as $\mathsf{BasicPub}^{hy}$ is IND-CCA secure.

  Thereby, $\mathcal{A}_3$ can not employ the decryption oracle of $\mathsf{BasicPub}^{hy}$ to answer decryption queries issued by $\mathcal{A}$ if the corresponding $coin_i = 1$.
- **Issue 2**. In the Challenge stage, $\mathcal{A}$ outputs ID_{ch} and two M_0, M_1 on which it wishes to be challenged. $\mathcal{A}_3$ gives its challenger M_0, M_1 as the messages that it wishes to be challenged on. The challenger gives $\mathcal{A}_3$ the ciphertext $C = \langle U, V, W\rangle = \langle rP, \sigma \oplus H_2(e(Q_{\mathsf{ID}}, P_{pub})^r), M_c \oplus H_4(\sigma)\rangle$ such that C is the encryption of M_c for random $c \in \{0,1\}$. Let $\langle \mathsf{ID}_{ch}, Q, b, coin\rangle$ be the corresponding tuple on the H_1^{list}. According to the above algorithm, if $coin = 0$ $\mathcal{A}_3$ aborts the game and the attack fails, otherwise $\mathcal{A}_3$ will modify C to be $C' = \langle U', V', W'\rangle = \langle b^{-1}U, V, W\rangle$ and relays C' to $\mathcal{A}$ as the challenge ciphertext. Boneh and Franklin claimed that C' is also a proper FullIdent encryption result of M_c under the public key $\mathsf{ID}_{ch} = Q = bQ_{\mathsf{ID}}$. However, if C' is a valid ciphertext of M_c in FullIdent, we have $r' = rb^{-1}$, $H_4(\sigma) = H_4(\sigma')$, $H_4(\sigma) = r$, $H_4(\sigma') = r'$. These facts imply that $b = 1$. Be aware of that b is randomly chosen from $\mathbb{Z}_q^*$, thereby the probability that C' is a valid ciphertext corresponding to M_c in FullIdent is $1 - 1/q$. On the other hand, for the same reason explained in Issue 1, $\mathcal{A}$ will reject the modified ciphertext with overwhelming probability.

Therefore, these two lapses render the Reduction 4 in the original proof invalid. By the way, Galindo only found out Issue 1 in [10].

3.2 Flipping coin technique

In the proof of Reduction 2, Boneh and Franklin borrowed the technique from Coron's analysis of the Full Domain Hash signature scheme [9], which we refer to it as flipping coin technique. More precisely, $\mathcal{A}_1$ answers H_1 queries according to the result of flipping a coin when simulating the H_1 random oracle for $\mathcal{A}_2$, i.e. before answering a new H_1-query at ID_i, $\mathcal{A}_2$ will generate a random $coin \in \{0,1\}$ with probability $\Pr[coin = 0] = \delta$,

- If $coin = 0$, return $Q_i = b_iP \in \mathbb{G}_1^*$.
- If $coin = 1$, return $Q_i = b_iQ_{\mathsf{ID}} \in \mathbb{G}_1^*$.

In Phase 1, only when $coin_i = 0$ could $\mathcal{A}_1$ answer the private key query $\langle \mathsf{ID}_i \rangle$ properly, because $Q_i = b_i P$ enables $\mathcal{A}_1$ to extract the private key as $d_i = b_i P_{pub}$. In the Challenge stage, only the case $coin_i = 1$ allows $\mathcal{A}_1$ to utilize $\mathcal{A}_2$'s guess to win the game, because $Q_i = b_i Q_{\mathsf{ID}}$ enables $\mathcal{A}_2$ to make use of the homomorphic relationship.

Waters [20] generalized such "flipping coin technique" as *Partitioning Reduction*: creating a reduction algorithm $\mathcal{B}$ that partitions the identity space into two parts (1) identities for which it can create private keys; (2) identities that it can use in the challenge phase. simulator hopes that the extraction/decryption queries and the challenge identity fall favorably in the partition, then the simulation is identical to the real attack in the adversary's view and the attack successes. The partition of identity space is only determined by the simulator (the distribution of $coin$) and independent to adversary's particular behavior, thus enables the possibility of perfect simulation computable.

When Boneh and Franklin applied this identical technique in Reduction 4, it does not work for both the challenger and the adversary $\mathcal{A}$ will check the validity of ciphertext, as pointed out in Issue 1 and Issue 2. The reason is that Fujisaki and Okamoto transformation [12] removes the homomorphic relationship between the ciphertext and its corresponding key, ill-formed ciphertext will be rejected with overwhelming probability, thereby the simulation fails.

4 Analysis of Galindo and Nishioka's Proofs

In this section we investigate the two subsequent revised proofs provided by Galindo and Nishioka, respectively.

4.1 Galindo's proof

Galindo [10] tried to fix the Reduction 4 by modifying the simulation method of random oracle H_1. His revised proof of Reduction 4 is shown as follows.

Setup. Same as BF-IBE's.

H_1-**queries**. Before initializing H_1^{list}, $\mathcal{A}_3$ selects a random $j \leftarrow \{1, \ldots, q_{H_1}\}$. When $\mathcal{A}$ queries H_1 at ID_i, $\mathcal{A}_3$ responds as follows: if $i \neq j$, it picks $b_i \leftarrow \mathbb{Z}_q^*$ and sets $Q_i = b_i P$, adds $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ to the list. If $i = j$, it sets $Q_i = Q_{\mathsf{ID}}$, adds $\langle \mathsf{ID}_i, Q_i, * \rangle$ to the list. Finally, $\mathcal{A}_3$ sends Q_i to $\mathcal{A}$.

Phase 1 - Extraction queries. When $\mathcal{A}$ asks for the private key of ID_i, $\mathcal{A}_3$ runs the above algorithm and gets $H_1(\mathsf{ID}_i) = Q_i$, where $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ is the corresponding entry in H_1^{list}. If $i = j$, then $\mathcal{A}_3$ aborts the game. Otherwise, it sets $d_i = b_i P_{pub}$. Finally, $\mathcal{A}_3$ gives d_i to $\mathcal{A}$.

Phase 1 - Decryption queries. $\mathcal{A}_3$ answers to decryption query $\langle \mathsf{ID}_i, C_i \rangle$ as follows. It runs H_1-queries algorithm and let $\langle \mathsf{ID}_i, Q_i, b_i \rangle \in H_1^{list}$. If $i \neq j$, then $\mathcal{A}_3$ retrieves the private key d_i and decrypts C_i using the decryption algorithm. If $i = j$, then $Q_i = Q_{\mathsf{ID}}$, and the decryption of $\langle \mathsf{ID}_i, C_i \rangle$ is the same as the decryption of C_i under $\mathsf{BasicPub}^{hy}$. Then, $\mathcal{A}_3$ asks its challenger to decrypt C_j and relays the answer to $\mathcal{A}$.

Challenge. $\mathcal{A}$ outputs a public key ID_{ch} and two messages M_0, M_1 on which it wishes to be challenged. $\mathcal{A}_3$ proceeds as follows. If $\mathsf{ID}_{ch} \neq \mathsf{ID}_j$, it aborts the game and the attack against $\mathsf{BasicPub}^{hy}$ failed. Otherwise, it sends M_0, M_1 to its own challenger and gets back C, the encryption of M_c for a random bit c under $\mathsf{BasicPub}^{hy}$. Finally, $\mathcal{A}_3$ relays C to $\mathcal{A}$, which is an also encryption of M_c under ID_{ch} for $\mathsf{FullIdent}$.

The **Phase 2** and **Guess** stage are identical to BF-IBE's.

In this game $\mathcal{A}_3$'s simulation can be aborted for two reasons: (1) in Phase 1 $\mathcal{A}$ issues the private key query of ID_j, or (2) in Challenge stage, the challenge identity $\mathsf{ID}_{ch} \neq \mathsf{ID}_j$. Note that $\mathcal{A}_3$ will not abort in Phase 2, since in this case $\mathcal{A}$ is not allowed to query the private key for $\mathsf{ID}_{ch} = \mathsf{ID}_j$.

Let $\mathcal{E}_1$ be the event that $\mathcal{A}_3$ aborts due to (1), $\mathcal{E}_2$ be the event that $\mathcal{A}_3$ aborts due to (2). The probability that $\mathcal{A}_3$ does not abort is $\Pr[\neg\mathcal{E}_1 \wedge \neg\mathcal{E}_2] = \Pr[\neg\mathcal{E}_2|\neg\mathcal{E}_1] \Pr[\neg\mathcal{E}_1]$.

Galindo deemed that the upper bound for $\Pr[\mathcal{E}_1]$ was q_E/q_{H_1}, since the maximum number of private extraction queries is q_E; the lower bound for $\Pr[\neg\mathcal{E}_2|\neg\mathcal{E}_1]$, that is the probability that $\mathcal{A}$ choose ID_j as the challenge identity, was $1/q_{H_1}$. Therefore, he concluded

$$\Pr[\mathcal{A}_3 \text{ does not abort}] \geq \frac{1}{q_{H_1}} \left(1 - \frac{q_E}{q_{H_1}}\right)$$

Now we point out three issues which may be overlooked in Galindo's proof.

- **Issue 1**. According to the definition of IND-ID-CCA game, q_{H_1} is unknown to the challenger $\mathcal{A}_3$ until the end of the game. So the execution of $\mathcal{A}_3$'s selecting a random $j \leftarrow \{1, \ldots, q_{H_1}\}$ at the beginning of simulation is questionable. In the other side, in order to provide a general and valid proof, the construction of $\mathcal{A}_3$ should be independent of the concrete behavior of adversary $\mathcal{A}$. In a word, this issue make the proof does not hold in a general sense.
- **Issue 2**. Even Issue 1 could be ignored, here follows issue 2. In the challenge stage, when $\mathcal{A}$ outputs the target identity ID_{ch}, the simulator need to judge if $\mathsf{ID}_{ch} = \mathsf{ID}_j$. In fact, the exact number of H_1 queries $\mathcal{A}$ in Phase 1 may differ from different adversaries in different games. Moreover, in some cases whether "ID_j" exists is unknown, thus the probability of "$\mathcal{A}_3$ does not abort" is immeasurable. For example, suppose the random $j = 10$ and an $\mathcal{A}$ issues only three H_1 queries in Phase 1, then the so called ID_j even does not exist. At least, it is fair to say the simulation algorithm is

not well defined, although the underlying idea may be correct.

- **Issue 3**. Even both Issue 1 and Issue 2 could be fixed, there is issue 3 following. The result of $\Pr[\neg\mathcal{E}_2|\neg\mathcal{E}_1] \geq 1/q_{H_1}$ is implied from the hypothesis that in the challenge stage the adversary $\mathcal{A}$ will randomly picks the target ID_{ch} from the current H_1^{list}. First, due to the same reason of Issue 2, whether the so called "ID_j" exists is a question, thereby the probability $\Pr[\mathsf{ID}_{ch} = \mathsf{ID}_j]$ is not well defined itself. Second, this goes against the definition of IND-ID-CCA which states that the target ID_{ch} can be chosen without any restriction, in particular outside the current H_1^{list}. Someone may argue that if the adversary $\mathcal{A}$ does not choose ID_{ch} from the current H_1^{list}, the advantage against the IND-ID-CCA game is statistically close to 0. Remember that $\mathcal{A}$ could issue the corresponding H_1 query in Phase 2. Besides, there is no evidence gurantees that the adversary $\mathcal{A}$ will choose the target identity uniformly from either inside or outside the current H_1^{list}.

From the above analyses, we think the revised proof proposed by Galindo is not infallible.

4.2 Nishioka's proof

In IndoCrypt 2005, Nishioka gave a new proof for the security of BF-IBE scheme in [17], claimed that it has a tighter security reduction than had been previously believed. Realizing that there are some problems in Galindo's proof, Nishioka claimed that Galindo's proof could be revised by his new proof. Unfortunately, we think the new proof share the similar fundamental problems as Galindo's proof.

Nishioka's proof for Reduction 4 is quite similar to Galindo's proof except three minor alterations. The first alteration is in the simulation of H_1-queries: in [10] $\mathcal{A}_3$ selects a random $j \in \{1, \ldots, 1 + q_{H_1}\}$, while in [17] $\mathcal{A}_3$ selects a random $j \in \{1, \ldots, 1 + q_{H_1} + q_D\}$. The second alteration is in [17] $\mathcal{A}_3$ maintains two lists named as H_1^{list1} and H_1^{list2}, where L_1 and L_2 are their corresponding size. H_1^{list1} is used to save $\mathcal{A}_3$'s responses to H_1-queries and decryption queries, while H_1^{list2} is used to save $\mathcal{A}_3$'s responses to extraction queries. The last alternation is replacing $\mathsf{ID}_i = \mathsf{ID}_j$ with $i = L_1 - 1$ in the challenge stage. The rest parts are identical to Galindo's proof.

In order to compute the probability that $\mathcal{A}_3$ does not abort during the simulation, Nishioka defines $\mathcal{E}_1$ as the event that $\mathcal{A}_3$ issues a private key query ID_j which corresponds to the tuple $\langle \mathsf{ID}_j, Q_{\mathsf{ID}}, * \rangle$ on H_1^{list1} during Phase 1 or 2, and defines $\mathcal{E}_2$ as the event that $\mathcal{A}$ sets the challenge identity ID_{ch} that does not correspond to the tuple $\langle \mathsf{ID}_j, Q_{\mathsf{ID}}, * \rangle$ on H_1^{list1}. Then Nishioka claimed that $\Pr[\mathcal{A}_3 \text{ does not abort}] = \Pr[\neg\mathcal{E}_1 \wedge \neg\mathcal{E}_2] = \Pr[\neg\mathcal{E}_2] \geq 1/(1 + q_{H_1} + q_D)$ (Equation. 2 in Section 3.2 in [17]). We summarize the lapses in Nishioka's proof as follows.

- **Issue 1**. The execution of $\mathcal{A}_3$'s selecting a random $j \leftarrow \{1, \ldots, 1 + q_{H_1} + q_D\}$ when initializing H_1^{list1} is doubtable. The reason is the same as Issue 1 of Galindo's proof.
- **Issue 2**. The computation of $\Pr[\mathcal{A}_3 \text{ does not abort}] = \Pr[\neg\mathcal{E}_1 \wedge \neg\mathcal{E}_2] = \Pr[\neg\mathcal{E}_2] \geq 1/(1 + q_{H_1} + q_D)$ is not correct, because the author missed to count in the contribution of event $\mathcal{E}_1$. The result should be corrected as $\Pr[\mathcal{A}_3 \text{ does not abort}] = \Pr[\neg\mathcal{E}_1 \wedge \neg\mathcal{E}_2] = \Pr[\neg\mathcal{E}_2|\neg\mathcal{E}_1]\Pr[\neg\mathcal{E}_1] \geq \frac{1}{1+q_{H_1}+q_D}\left(1 - \frac{q_E}{1+q_{H_1}+q_D}\right)$

These two issues render Nishioka's proof for Reduction 4 not intact.

4.3 How to fix their Proofs

The main idea of proving Reduction 4 is suppose there is an IND-ID-CCA adversary $\mathcal{A}$ against FullIdent with advantage $\epsilon(k)$, then create an IND-CCA adversary $\mathcal{A}_3$ against $\mathsf{BasicPub}^{hy}$ with the help of $\mathcal{A}$. The validness of Reduction 4 is determined by the possibility of $\mathcal{A}_3$ does not abort during simulation. Galindo and Nishioka dumped the "flipping coin technique" by pre-specifying a index at the beginning of the game. Their constructions of simulate algorithm are related to the concrete behavior of distinct adversaries, which make their proof lose generality.

We summarize the guidelines that a valid IND-ID-CCA proof should follows.

- The construction of the simulator should be general. More exactly, the simulation algorithm should be independent of the adversary's particular behavior, such as the exact number H_1 queries, extraction/decryption queries a adversary makes in Phase 1 and Phase 2. Otherwise the reduction is not a general one.
- The adversary must be handled strictly according to the definition of the IND-ID-CCA game, no extra hypothesis should be imposed on it, such as (1) Which target identity it will choose in the challenge stage? (2) How does the adversary choose the target identity? From which set and the choices comply to what probabilistic distribution?
- The simulation algorithm must ensure the probability of perfect simulation to be computable, thereby the advantage of the simulator against the underlying problem is measurable. Otherwise the security reduction is meaningless.

How to fix their proofs? Note that the adversary is modeled as an polynomial time algorithm, thereby the number

of H_1 queries, extraction queries and decryption queries are parameterized by the security parameter k. Thus we can simply fix Issue 1 of Galindo and Nishioka's proof by substituting q_{H_1} or $1 + q_{H_1} + q_D$ with a sufficiently large number N to initialize H^{list} (e.g. set N as the upper bound for the number of total queries an adversary issues during the IND-ID-CCA game, according to common practice $N = 2^{60}$ is a natural choice). Issue 2 and Issue 3 of Galindo's proof could be fixed by well defining the simulation algorithm, i.e. using the current size of the H^{list} to replace a particular ID_i, thereby make the simulation algorithm well defined. Issue 2 of Nishioka's proof has been corrected in the above related analysis.

5 IND-sID-CCA security of BF-IBE

In this section, we provide a security proof for BF-IBE in selective-ID model, then try to achieve fully security based on it.

Theorem 5.1. *Let H_1 be a random oracle. Then* FullIdent *is* IND-sID-CCA *secure assuming the CBDH assumption is hard in groups generated by $\mathcal{G}$. Concretely, suppose there is an* IND-sID-CCA *adversary $\mathcal{A}$ that has advantage $\epsilon(k)$ against the* FullIdent*. Then there is an* IND-CCA *adversary $\mathcal{A}_3$ that has advantage $\epsilon(k)$ against* $\mathsf{BasicPub}^{hy}$*. Its running time is $O(time(\mathcal{A}))$.*

Proof. We construct an IND-CCA adversary $\mathcal{A}_3$ that uses $\mathcal{A}$ to gain advantage against $\mathsf{BasicPub}^{hy}$. The game starts with the challenger first generates the public key $K_{pub} = \langle q, \mathbb{G}_1, \mathbb{G}_2, e, n, P, P_{pub}, Q_{\mathsf{ID}}, H_2, H_3, H_4 \rangle$ and a private key $d_{\mathsf{ID}} = sQ_{\mathsf{ID}}$. The challenger gives K_{pub} to algorithm $\mathcal{A}_3$. $\mathcal{A}_3$ mounts an IND-CCA attack on the the key K_{pub} using the help of algorithm $\mathcal{A}$. $\mathcal{A}_3$ interacts with $\mathcal{A}$ as follows.

Init. $\mathcal{A}$ outputs an identity ID_{ch} where it wishes to be challenged.

Setup. Same as BF-IBE's.

H_1-**queries**. To respond to H_1 queries, $\mathcal{A}_3$ maintains a list of tuples $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ which is referred as H_1^{list}. The list is initially empty. When $\mathcal{A}$ queries H_1 at a point ID_i, $\mathcal{A}_3$ responds as follows:

1. If the query ID_i already appears on the H_1^{list} in a tuple $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ then $\mathcal{A}_3$ responds with $H_1(\mathsf{ID}_i) = Q_i$.
2. Otherwise, if $\mathsf{ID}_i = \mathsf{ID}_{ch}$, $\mathcal{A}_3$ sets $b_i = *$ and $Q_i = Q_{\mathsf{ID}}$; else $\mathcal{A}_3$ generates a random $b_i \in \mathbb{Z}_q^*$ and computes $Q_i = b_i P$.
3. $\mathcal{A}_3$ adds the tuple $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ to H_1^{list} and responds to $\mathcal{A}$ with $H_1(\mathsf{ID}_i) = Q_i$.

Phase 1 - Extraction queries. When $\mathcal{A}$ asks for the private key associate to ID_i, $\mathcal{A}_3$ runs the above algorithm and gets $H_1(\mathsf{ID}_i) = Q_i$, where $\langle \mathsf{ID}_i, Q_i, b_i \rangle$ is the corresponding entry in H_1^{list}. Observing that $Q_i = b_i P$, therefore the corresponding private key is $d_i = b_i P_{pub}$. Finally, $\mathcal{A}_3$ gives d_i to $\mathcal{A}$. The request $\langle \mathsf{ID}_{ch} \rangle$ will be denied.

Phase 1 - Decryption queries. Let $\langle \mathsf{ID}_i, C_i \rangle$ be a decryption query issued by algorithm $\mathcal{A}$. Let $C_i = \langle U_i, V_i, W_i \rangle$. When $\mathsf{ID}_i \neq \mathsf{ID}_{ch}$, $\mathcal{A}_3$ runs H_1-queries algorithm and let $\langle \mathsf{ID}_i, Q_i, b_i \rangle \in H_1^{list}$, then retrieves the private key d_i and decrypts C_i using the decryption algorithm. If $\mathsf{ID}_i = \mathsf{ID}_{ch}$, $\mathcal{A}_3$ relays the decryption query with the ciphertext $\langle U_i, V_i, W_i \rangle$ to the challenger and relays the challenger's response back to $\mathcal{A}$.

Challenge. Once $\mathcal{A}$ decides that Phase 1 is over and outputs two messages M_0, M_1 which it wishes to be challenged on. $\mathcal{A}_3$ responds as follows: first $\mathcal{A}_3$ gives its challenger the message M_0, M_1. The challenger responds with a $\mathsf{BasicPub}^{hy}$ ciphertext $C = \langle U, V, W \rangle$ such that C is the encryption of M_c for a random $c \in \{0, 1\}$. Next, $\mathcal{A}_3$ responds to $\mathcal{A}$ with the challenge C.

Phase 2 - Private key queries. $\mathcal{A}_3$ responds to the extraction queries in the same way as it did in Phase 1.

Phase 2 - Decryption queries. $\mathcal{A}_3$ responds to the decryption queries in the same way as it did in Phase 1 except that $\langle \mathsf{ID}_i, C_i \rangle = \langle \mathsf{ID}_{ch}, C \rangle$ is denied.

Guess. Eventually, adversary $\mathcal{A}$ outputs a guess c' for c. Algorithm $\mathcal{A}_3$ outputs c' as its guess for c.

All the responses to H_1-queries are as in real attack since each response is uniformly and independently in $\mathbb{G}_1^*$. All the responses to private key extraction queries and decryption queries are valid. Algorithm $\mathcal{A}_3$ wouldn't abort during the simulation because $\mathcal{A}$'s view is identical to its view in the real attack. By definition of algorithm $\mathcal{A}$, we have that $\left|\Pr[c = c'] - \frac{1}{2}\right| \geq \epsilon(k)$. Note that $\Pr[\mathcal{A}_3 \text{does not abort}] = 1$, this shows that $\mathcal{A}_3$'s advantage against $\mathsf{BasicPub}^{hy}$ is at least $\epsilon(k)$ as required. □

In the selective model, the identity space can always be "tightly" partitioned, so the reduction is tighter linked to the chosen ciphertext security of $\mathsf{BasicPub}^{hy}$.

5.1 Selective secure implies fully secure

In [2] Boneh and Boyen proved the following theorem which quantifies the relationship between selective-ID IBE security and fully IBE security in the random oracle model.

Theorem 5.2. *Let $\mathcal{E}$ be a selective-ID secure IBE. Suppose identities in $\mathcal{E}$ are n-bits long. Let H be a hash function $H : \{0,1\}^* \to \{0,1\}^n$ modeled as a random oracle. Then $\mathcal{E}_H$ is a fully secure IBE in the random oracle model. The reduction factor is $1/N$, where N is the maximum number of oracle calls to H that the adversary can make.*

As to BF-IBE scheme, if we first hash arbitrary identities in $\{0,1\}^*$ to binary strings of length n using a collision resistant function with n-bits output, such as SHA-1 whose output is 160 bits, then it is easy to demonstrate BF-IBE is fully secure according to Theorem 5.1 and Theorem 5.2.

6 Conclusion

In this paper, we identify some defects in the previous proofs of BF-IBE. While the original proof [3] fails for the manipulated ciphertext will be rejected by the decryption oracle, the subsequent proofs [10] [17] which intended to fix the problem also have some questionable issues. We show that Galindo and Nishioka's proofs could be mended by maximizing the number of queries. We also present another proof by confining the identity space into a finite set.

It is fair to say the aforementioned proofs are valid in the practical view, but far from perfect in theoretical view for two reasons: (1) both of them were proved via Reduction 4, Reduction 3 and Reduction 1, which happens to be the reason make the security reduction looser; (2) the simulation algorithms are implicitly depicted by a heuristic upper bound N for the number of queries the adversary can make. For this reason, we think how to give a elegant and tighter proof for BF-IBE scheme in the full IBE security model under the CBDH assumption is still a interesting problem. Our future work is reducing the CCA security of BF-IBE directly to some underlying assumption without the intermediate reductions.

References

[1] S. S. Alriyami, K. G. Paterson, and R. Holloway. Certificateless public key cryptography. *Advances in Cryptology - Asiacrypt 2003*, 2894:452–473, 2003.

[2] D. Boneh and X. Boyen. Efficient selective-id secure identity based encryption without random oracles. *Proceedings of Eurocrypt 2004*, 3027:223–238, 2004.

[3] D. Boneh and M. K. Franklin. Identity-based encryption from the weil pairing. *Advances in Cryptology - CRYPTO 2001*, 2139:213–229, 2001.

[4] D. Boneh and M. K. Franklin. Identity-based encryption from the weil pairing. *SIAM J. of Computing*, 32:586–615, 2003.

[5] X. Boyen. Multipurpose identity-based signcryption - a swiss army knife for identity-based cryptography. *Advances in Cryptology - CRYPTO 2003*, 2729:383–399, 2003.

[6] R. Canetti, S. Halevi, and J. Katz. A forward-secure public-key encryption scheme. *Advances in Cryptology - Eurocrypt 2003*, 2656:255–271, 2003.

[7] R. Canetti, S. Halevi, and J. Katz. Chosen-ciphertext security from identitybased encryption. *Advances in Cryptology - Eurocrypt 2004*, 3027:207–222, 2004.

[8] Clifford Cocks. An Indentity Based Encryption Scheme Based on Quadratic Residues. *Institute of Mathematics and Its Applications International Conference on Cryptigraphy and Coding - Proceedinds of IMA 2001*, 2260:360–363, 2001.

[9] J.-S. Coron. On the exact security of full domain hash. *CRYPTO 2000*, 1880:229–235, 2000.

[10] David Galindo. Boneh-franklin identity based encryption revisited. *Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005 Proceedings*, 3580:791–802, 2005.

[11] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. pages 537–554, 1999.

[12] E. Fujisaki and T. Okamoto. Secure integration of asymmetric and symmetric encryption schemes. *CRYPTO 1999: Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pages 537–554, 1999.

[13] C. Gentry. Certificate-based encryption and the certificate revocation problem. 2656:272–293, 2003.

[14] C. Gentry and A. Silverberg. Hierarchical id-based cryptography. *Advances in Cryptology - ASIACRYPT 2002*, 2501:548–566, 2002.

[15] J. Horwitz and B. Lynn. Toward hierarchical identity-based encryption. *Advances in Cryptology - Eurocrypt 2002*, 2322, 2002.

[16] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. pages 62–73, 1995.

[17] M. Nishioka. Reconsideration on the security of the boneh-franklin identity-based encryption scheme. *Progress in Cryptology - INDOCRYPT 2005, 6th International Conference on Cryptology in India*, 3797:270–282, 2005.

[18] M. K. Ryuichi Sakai, Kiyoshi Ohgishi. Cryptosystems based on pairing. *The 2001 Symposium on Cryptography and Information Security, Japan*, 45:26–28, 2001.

[19] A. Shamir. Identity-based cryptosystems and signatures schemes. *Advances in Cryptology - Crypto 1984*, 196:47–53, 1984.

[20] B. Waters. Dual system encryption: Realizing fully secure ibe and hibe under simple assumptions. *CRYPTO '09: Proceedings of the 29th Annual International Cryptology Conference on Advances in Cryptology*, pages 619–636, 2009.