Fuzzy Identity Based Encryption

Preliminary Version

Amit Sahai* Brent R. Waters

Abstract

We introduce a new type of Identity Based Encryption (IBE) scheme that we call Fuzzy Identity Based Encryption. A Fuzzy IBE scheme allows for a private key for an identity, I, to decrypt a ciphertext encrypted with another identity, I', if and only if the identities I and I' are close to each other as measured by some metric such as Hamming distance or set overlap. A Fuzzy IBE scheme can be applied to enable encryption using biometric measurements as identities. The error-tolerance of a Fuzzy IBE scheme allows for the use of biometric identities, which inherently will have some noise each time they are sampled.

In this paper we present a construction of a Fuzzy IBE scheme that uses groups with efficiently computable bilinear maps. Additionally, our construction does not use Random Oracles. We prove the security of our scheme under the Selective-ID security model.

1 Introduction

There has been recent interest about the challenge of generating cryptographic keys from biometric inputs. The primary difficulty in generating a strong key from a biometric input is that the measured value of a biometric can change slightly upon each sampling. This effect can be explained by differences in sampling devices, environmental noise, or small changes in the human trait itself. This inherent non-determinism makes it difficult to extract a cryptographic key from a biometric input.

Recent work has produced techniques to derive cryptographic keys from biometric inputs for symmetric key applications. For example, Monrose et al. [14, 13, 12] develop techniques to extract secrets from keyboard typing dynamics and, later, voice prints by using a form of error-tolerant secret sharing. Other work by Davida *et al.* [4] and Juels and Wattenberg [9] use error-correcting codes to compensate for the noise in the biometric input.

These techniques are useful for symmetric key cryptography applications such as password authentication and symmetric key encryption. However, there does not seem to be a clear way to move these techniques into the realm of public key cryptography. In particular the work above seems does not fit into the paradigm of Identity-Based Encryption.

We propose a new type of Identity-Based Encryption that we call *Fuzzy Identity-Based Encryption*. In a Fuzzy Identity-Based Encryption scheme a user with secret key for the identity I is able to decrypt a ciphertext encrypted with the public key I' if and only if I and I' are within a certain distance of each other as judged by some metric. For the remainder of this paper we view identities as a set of n elements drawn from some large Universe of elements. We use the set overlap between two identities to measure their similarity. (If we restrict ourselves to equal size sets we could view the set difference as the distance between two identities.) In general, a Fuzzy Identity-Based Encryption scheme could be built from other distance metrics.

Motivating Example The existence of a practical Fuzzy Identity-Based Encryption scheme would allow for the encryption of data using a biometric input as the public key. This could be useful in the following scenario. Consider a patient that is rushed into an emergency medical visit. A collection of medical data will be created from the visit including possibly the results of

*Amit Sahai's research was supported by generous grants from the NSF ITR program, as well as a Sloan Foundation Fellowship.

tests that will not be ready until days later. The patient would like to be able to conveniently access this data later by retrieving it from a public storage server. However, since the privacy of medical information is considered to be very important, the data would need to be encrypted in such a way that only the patient could access it.

Given the urgency of the initial visit it is unrealistic to require that the patient has on his person a device that can store a cryptographic key or to require that he remember a unique identifier (as might be required if we were to use Identity-Based Encryption in the traditional manner). However, the medical staff would be able to measure an inherent biometric identity from the patient. Using a Fuzzy IBE scheme the medical staff could use this biometric identity to encrypt information for the patient so that he could later retrieve at his convenience.

Advantages of biometric-based IBE over traditional IBE In many situations, using biometric-based IBE would have a number of important advantages over "traditional" IBE. We argue that the use of biometric identities fits the framework of Identity-Based Encryption very well and is a very valuable application of it.

First, the process of obtaining a secret key from an authority is very natural and straightforward. In traditional Identity-Based Encryption schemes a user with a certain identity, for example "Bob Smith", will need to go to an authority to obtain the private key corresponding to the identity. In this process the user will need to "prove" to the authority that he is indeed entitled to this identity. This will typically involve presenting supplementary documents or credentials. The type of authentication that is necessary is not always clear and robustness of this process is questionable (the supplementary documents themselves could be subject to forgery) Typically, there will exist a tradeoff between a system that is expensive in this step and one that is less reliable.

In contrast, if a biometric is used as an identity then the verification process for an identity is very clear. The user must demonstrate ownership of the biometric under the supervision of a well trained operator. If the operator is able to detect imitation attacks, for example playing the recording of a voice, then the security of this phase is only limited by the quality of the biometric technique itself. We emphasize that the biometric measurement for an individual need not be kept secret (indeed it won't be if it is to be used as a public key). We must only guarantee that an attacker cannot fool the human attendee that delegates private keys into believing that an attacker owns a biometric identity that he does not.

Secondly, using a biometric as an identity has the advantage that identities are unique if the underlying biometric is of a good quality. Some types of traditional identities such as the name "Bob Smith" will clearly not be unique and this could lead to problems, although Hierarchial Identity-Based Encryption [6, 8] does alleviate this problem.

Finally, the fact that a biometric identity is an inherent trait and therefore is always with a person is useful. In several situations such as our medical example the user might not have the foresight to carry an cryptographic devices or even obtain a unique traditional identity.

Our Contributions We formalize the notion of Fuzzy Identity-Based Encryption and provide a construction for a Fuzzy Identity-Based Encryption scheme. Our construction uses groups for which an efficient bilinear map exists, but for which the Computational Diffie-Hellman problem is assumed to be hard. We achieve our result by applying the techniques of Shamir Secret Sharing [17] where a polynomial can be reconstructed in the exponent of a group.

We prove our scheme secure under an adapted version of the Selective-ID security model first proposed by Canetti et al. [3]. Additionally, our construction does not use Random Oracles. We reduce the security of our scheme to an assumption that is similar to the Bilinear Decisional Diffie-Hellman assumption.

1.1 Related Work

Shamir [18] first proposed the concept of Identity-Based Encryption. Boneh and Franklin [2] presented the first Identity-Based Encryption scheme that was both practical and secure. Their solution also made use of groups for which there was an efficiently computable bilinear map, but for which the Computational Diffie-Hellman problem is believed to be hard.

Canetti et al. [3] proposed the first construction for IBE that was provably secure outside the Random Oracle model. To prove Security they described a slightly weaker model of security known as the Selective-ID model, in which the adversary declares which identity he will attack before the global public parameters are generated. Since their construction views identities as bit strings a Bilinear map computation is required for each bit in the identity.

Boneh and Boyen [1] describe two IBE scheme that are also proven to be secure in the Selective-ID model. Additionally, their two constructions only require respectively two and one pairing computations per decryption and are thus significantly more efficient.

Our work can be viewed as a new method for a standard IBE scheme that is secure without Random Oracles in the Selective-ID model. Our Fuzzy IBE scheme becomes a standard one when the error-tolerance parameter, d, is set to 0. However, since the decryption phase requires a pairing computation for every bit of the identity it would serve as a much less efficient standard IBE scheme than that of Boneh and Boyen [1].

Other work in applying biometrics to cryptography has focused on the derivation of a secret from a biometric. This secret can be then used for operations such as encryption or UNIX style password authentication.

Monrose et al. [14, 13, 12] examined the use of keystroke dynamics and later voice prints as biometrics for this purpose. The authors use a type of secret sharing in which the biometric would decided which keyshares from a stored set to use in reconstructing the secret. Some stored shares are bogus to prevent someone with the wrong biometric from unlocking the secret. Their scheme provides error-tolerance by storing two valid secrets for points in the biometric measurement that are not reliable for a particular user.

Davida et al. [4] proposed a scheme where they store the check bits for an error correcting code of a biometric input along with the hash of the input. The check bits are used to compute the original input from one that is reasonably close. Juels and Wattenberg [9] improve upon this work using a novel technique in which the biometric input is treated as a corrupted codeword.

Juels and Sudan [10] presented a scheme that uses set overlap as the measurement between biometric templates. Dodis, Rezyin, and Smith [5] describe a general primitive for extracting uniform randomness from biometric inputs. Additionally, they give constructions for the metrics of Hamming distance, set overlap, and edit distance.

The primary distinguishing feature of our work from the other related work on biometrics above is that we view the biometric input as public identity as opposed to a secret that the human poses. Our only physical requirement is that the biometric cannot be imitated such that a trained human operator would be fooled.

Finally, the use of polynomial interpolation within the exponents was discussed with Philippe Golle and Jessica Staddon [7] during work on the problem of conjunctive keyword search on encrypted data.

1.2 Organization

The rest of the paper is organized as follows. In Section 2 we formally define a Fuzzy Identity-Based Encryption scheme including the Selective-ID security model for one. In Section 3 we describe the security assumption our scheme reduces to. In Section 4 we describe our construction of a Fuzzy Identity-Based Encryption scheme. In Section 5 we prove the security of our scheme. Finally, we conclude in Section 6.

2 Definitions

In this section we define two Selective-ID models of security for Fuzzy Identity Based Encryption. The Basic Fuzzy Selective-ID game is very similar to the standard Selective-ID model for Identity-Based Encryption with the exception that the adversary is only allowed to query for secret keys for identities which have less than d overlap with the target identity.

In the second game, which we call Multiple Fuzzy Selective-ID the adversary chooses several identities that are simultaneously targets. This model of security captures how an attack against a biometric identity will occur in a practice. The attacker will want to target a certain biometric identity, but the identities with which the ciphertexts are encrypted will deviate slightly from

the original identity. These deviations in practice will be from environmental noise and therefore will be independent of the public parameters. This game captures this case since the attacker can model the random distribution of identities in the multiple identities that he chooses to attack.

We describe the two security games below and show that the security of the Multiple Fuzzy Selective-ID game reduces to the Basic Fuzzy Selective-ID game in Section A of the Appendix.

Basic Fuzzy Selective-ID

Init The adversary declares the identity α that he wishes to be challenged upon.

Setup The challenger runs the setup phase of the algorithm and tells the adversary the public parameters.

Phase 1 The adversary is allowed to issue queries for private keys for an identity γ where $|\gamma \cap \alpha| < d$.

Challenge The adversary submits two equal length messages m_0, m_1. The challenger flips a random coin b and encrypts m_b with α. The ciphertext is passed to the adversary.

Phase 2 Phase 1 is repeated.

Guess The adversary outputs a guess b' of b.

The advantage of an adversary $\mathcal{A}$ in this game is defined as $\Pr[b' = b] - \frac{1}{2}$.

Definition 1 (Basic Fuzzy Selective-ID). *A scheme is secure in the Basic Fuzzy Selective-ID model of security if all computationally bound adversaries have at most a negligible advantage in the above game.*

Multiple Fuzzy Selective-ID

Init The adversary declares l identities $\alpha_1, \ldots, \alpha_l$ that he wishes to be challenged upon where l is bounded by a polynomial of the security parameter κ.

Setup The challenger runs the setup phase of the algorithm and tells the adversary the public parameters.

Phase 1 The adversary is allowed to issue queries for private keys for an identity γ where $|\gamma \cap \alpha_i| < d$ for all i.

Challenge The adversary submits $2l$ messages $m_{1,0}, \ldots, m_{l,0}$ and $m_{1,1}, \ldots, m_{l,1}$ where $m_{i,0}, m_{i,1}$ are of equal length for all i. The challenger flips a random coin b and encrypts $m_{1,b}, \ldots, m_{l,b}$ with the respective identities $\alpha_1, \ldots, \alpha_l$. The ciphertexts are passed to the adversary.

Phase 2 Phase 1 is repeated.

Guess The adversary outputs a guess b' of b.

The advantage of an adversary $\mathcal{A}$ in this game is $\Pr[b' = b] - \frac{1}{2}$.

3 Complexity Assumption

Let $\mathcal{G}_1$ be a group of prime order p with an admissible bilinear map, e, into $\mathcal{G}_2$ and g be a generator of $\mathcal{G}_1$. Our assumption follows.

Definition 2 (Decisional Bilinear Diffie-Hellman (BLDH) Assumption). *Suppose a challenger chooses $a, b, c, z \in \mathbf{Z}_p$ at random. The Decisional BLDH assumption is that no adversary is to be able to distinguish the tuple*

$$(A = g^a, B = g^b, C = g^c, Z = e(g, g)^{abc})$$

from the tuple

$$(A = g^a, B = g^b, C = g^c, Z = e(g, g)^z)$$

with more than a negligible advantage.

4 Our Construction

4.1 Description

Let $\mathcal{G}_1$ be bilinear group of prime order p and g be a generator of $\mathcal{G}_1$. Additionally, let $e : \mathcal{G}_1 \times \mathcal{G}_1 \rightarrow \mathcal{G}_2$ denote the bilinear map. A security parameter, κ, will determine the size of the groups.

We also define the Lagrange coefficient $\triangle_{i,S}$ for $i \in \mathbf{Z}_p$ and a set, S, of elements in $\mathbf{Z}_p$:

$$\triangle_{i,S}(x) = \prod_{j \in S, j \neq i} \frac{x-j}{i-j}.$$

Identities will be n element sets where the elements are members of $\mathbf{Z}_p^*$. Alternatively, we can describe an identity as a collection of n strings of arbitrary length and use a collision resistant hash function, H, to hash strings into members of $\mathbf{Z}_p^*$. Our construction follows:

Setup(n, d) First, choose $g_1 = g^y, g_2 \ \in \mathcal{G}_1$.

Next, choose $t_1, \ldots, t_n$ uniformly at random from $\mathcal{G}_1$. Let N be the set $\{1, \ldots, n\}$ and we define a function, T, as:

$$T(x) = g_2^{x^n} \prod_{i=1}^{n} t_i^{\triangle_{i,N}(x)}.$$

We can view T as the function $g_2^{x^n} g^{\omega(x)}$ for some $n-1$ degree polynomial ω.

The public key is published as: $g_1, g_2, t_1, \ldots, t_n$ and the private key is g_2^y.

Key Generation To generate a private key for identity I the following steps are taken. A $d-1$ degree polynomial q is randomly chosen such that $q(0) = y$. The private key will consist of two sets. The first set, $D = \{D_i\} \forall i \in I$, where the elements are constructed as

$$D_i = g_2^{q(i)} T(i)^{r_i},$$

where r_i is a random member of $\mathbf{Z}_p$ defined for all $i \in I$.

The other set is $d = \{d_i\} \forall i \in I$ where the elements are constructed as

$$d_i = g^{r_i}.$$

Encryption Encryption with the public key I' and message $m \in \mathcal{G}_2$ proceeds as follows.

First, a random value $s \in \mathbf{Z}_p$ is chosen. The ciphertext is then published as:

$$E = (I', E' = me(g_1, g_2)^s, E'' = g^s, \{E_i = T(i)^s\} \forall i \in I').$$

Decryption Suppose that a ciphertext, E, is encrypted with a key for identity I' and we have a key for identity I, where $|I \cap I'| \geq d$. Choose an arbitrary d-element subset, S, of $I \cap I'$.

Then, the ciphertext can be decrypted as:

$$\begin{aligned}
M &= E' \prod_{i \in S} \left(\frac{e(d_i, E_i)}{e(D_i, E'')} \right)^{\triangle_{i,S}(0)} \\
&= me(g_1, g_2)^s \prod_{i \in I \cap I'} \left(\frac{e(g^{r_i}, T(i)^s)}{e(g_2^{q(i)} T(i)^{r_i}, g^s)} \right)^{\triangle_{i,S}(0)} \\
&= me(g, g_2)^{ys} \prod_{i \in S} \frac{1}{e(g, g_2)^{q(i) s \triangle_{i,S}(0)}} \\
&= m.
\end{aligned}$$

The last equality is derived from using polynomial interpolation in the exponents. Since, the polynomial $sq(x)$ is of degree $d-1$ it can be interpolated using d points.

4.2 Computational Cost

Encryption of a message will consist of $n + 2$ exponentiations in the group $\mathcal{G}_1$. The cost of decryption will be dominated by $2d$ bilinear map computations.

4.3 Flexible Error-Tolerance

In our scheme the error-tolerance is set to a value d. However, in practice a party constructing a ciphertext might want more flexibility. For example, if a biometric input device happens to be less reliable it might be desirable to relax the set overlap parameters.

In practice this can be easily done. In addition to the regular attributes for an identity the private key dealer can consider all identities "default-1", "default-2", ...; if the one wants to encrypt a message with error-tolerance $d' = d - j$ then one can just make the encrypt under the identity plus the attributes "default-1" through "default-j".

5 Proof of Security

We prove that the security of our scheme in the Selective-ID model reduces to the hardness of the Decisional BLDH assumption.

Lemma 1. *If an adversary can break our scheme in the Basic Fuzzy Selective ID Model, then a simulator can be constructed to play the Modified BLDH game with a non-negligible advantage.*

Proof. Suppose there exists a computationally efficient adversary, $\mathcal{A}$, that can attack our scheme in the Selective-ID model with advantage ϵ. We build a simulator $\mathcal{B}$ that can play the Decisional BLDH game with advantage $\frac{\epsilon}{2}$.

The simulation proceeds as follows:

We first let the challenger set the groups $\mathcal{G}_1$ and $\mathcal{G}_2$ with an efficient bilinear map, e and generator g. The challenger flips a fair binary coin μ outside of $\mathcal{B}$'s view. If $\mu = 0$, the challenger sets $(A, B, C, Z) = (g^a, g^b, g^c, e(g, g)^{abc})$; otherwise it sets $(A, B, C, Z) = (g^a, g^b, g^c, e(g, g)^z)$ for random a, b, c, z.

Setup $\mathcal{B}$ will run $\mathcal{A}$ and receive the challenge identity, α, an n element set of members of $\mathbf{Z}_p$.

The simulator assigns the public parameters $g_1 = A$ and $g_2 = B$. It then chooses a random $n - 1$ degree polynomial $f(x)$ and calculates an $n - 1$ degree polynomial $u(x)$ such that $u(x) = -x^n$ for all $x \in \alpha$. Then, for i from 1 to n the simulator sets $t_i = g_1^{u(x)} g^{f(x)}$. Note that since $f(x)$ is a random $n - 1$ degree polynomial all t_i will be chosen at random from the adversaries view as in the construction.

Phase 1 $\mathcal{A}$ makes requests for private keys where the identity set overlap between the identities for the requested keys and α is less than d.

Suppose $\mathcal{A}$ asks a private key γ. We first define three sets of indices $\Gamma, \Gamma', S \subseteq \{1 \dots n\}$ in the following manner:

$$\Gamma = \{\gamma \cap \alpha\},$$

$$\Gamma' \text{ be any set such that } \Gamma \subseteq \Gamma' \subseteq \gamma \text{ and } |\Gamma'| = d - 1, \text{and}$$

$$S = \Gamma' \cup \{0\}.$$

Next, we define the decryption key components D_i and d_i for $i \in \Gamma'$ as:

$$D_i = g_2^{\lambda_i} T(i)^{r_i} \text{ where } s_i \text{ is chosen randomly in } \mathbf{Z}_p$$

and we let

$$d_i = g^{r_i}.$$

The intuition behind these assignments is that we are implicitly choosing a random $d - 1$ degree polynomial $q(x)$ by choosing its value for the $d - 1$ points randomly in addition to having $q(x_0) = a$.

The simulator also needs to calculate the decryption key values for all $i \in \gamma - \Gamma'$. To produce these we use an algebraic method from Boneh and Boyen [1]. The key components are calculated as:

$$D_i = (\prod_{j \in \Gamma'} g_2^{\lambda_j \triangle_{j,S}(i)}) \left(g_1^{\frac{-f(i)}{i^n + u(i)}} (g_2^{u(i)} g^{f(i)})^{r'_i} \right)^{\triangle_{0,S}(i)}$$

and

$$d_i = (g_1^{\frac{-1}{i^n + u(i)}} g^{r'_i})^{\triangle_{0,S}(i)}.$$

The value $i^n + u(i)$ will be non-zero for all $i \notin \alpha$,which includes all $i \in \gamma - \Gamma'$. This follows from the fact that $u(x)$ is an $n-1$ degree polynomial and that it agrees with the n degree polynomial $-x^n$ at n values of $x \in \alpha$. If it were to agree with $-x^n$ at any more points then it would have to have degree greater than $n-1$.

To show that these are valid keys let $r_i = (r'_i - \frac{y}{i^n + u(i)}) \triangle_{0,S}(i)$ and let $q(x)$ be the $d-1$ degree polynomial for which $q(0) = a$ and $q(i) = \lambda_i \forall i \in \Gamma'$. We then have:

$$\begin{aligned}
D_i &= (\prod_{j \in \Gamma'} g_2^{\lambda_j \triangle_{j,S}(i)}) \left(g_1^{\frac{-f(i)}{i^n + u(i)}} (g_2^{i^n + u(i)} g^{f(i)})^{r'_i} \right)^{\triangle_{0,S}(i)} \\
&= (\prod_{j \in \Gamma'} g_2^{\lambda_j \triangle_{j,S}(i)}) \left(g_2^a (g_2^{i^n + u(i)} g^{f(i)})^{\frac{-a}{i^n + u(i)}} (g_2^{i^n + u(i)} g^{f(i)})^{r'_i} \right)^{\triangle_{0,S}(i)} \\
&= (\prod_{j \in \Gamma'} g_2^{\lambda_j \triangle_{j,S}(i)}) \left(g_2^a (g_2^{i^n + u(i)} g^{f(i)})^{r'_i - \frac{a}{i^n + u(i)}} \right)^{\triangle_{0,S}(i)} \\
&= (\prod_{j \in \Gamma'} g_2^{\lambda_j \triangle_{j,S}(i)}) g_2^{a \triangle_{0,S}(i)} (T(i))^{r_i} \\
&= g_2^{q(x)} T(i)^{r_i}
\end{aligned}$$

Additionally, we have:

$$\begin{aligned}
d_i &= (g_1^{\frac{-1}{i^n + u(i)}} g^{r'_i})^{\triangle_{0,S}(i)} \\
&= (g^{r'_i - \frac{y}{i^n + u(i)}})^{\triangle_{0,S}(i)} \\
&= g^{r_i}
\end{aligned}$$

Therefore, the simulator is able to construct a private key for the identity γ. Furthermore, the distribution of the private key for γ is identical to that of the original scheme.

Challenge The adversary, $\mathcal{A}$, will submit two challenge messages m_1 and m_0 to the simulator. The simulator flips a fair binary coin ν and returns an encryption of m_ν. The ciphertext is output as:

$$E = (\alpha, E' = m_\nu Z, E'' = C, \{E_i = C^{f(i)}\} \forall i \in \alpha).$$

If $\mu = 0$, then $Z = g^{abc}$. Then the ciphertext is:

$$E = (\alpha, E' = m_\nu e(g,g)^{abc}, E'' = g^c, \{E_i = (g^c)^{f(i)} = T(i)^c\} \forall i \in \alpha).$$

This is a vaild ciphertext for the message m_ν under the identity α.

Otherwise, if $\mu = 1$, then $Z = g^z$. We then have $E' = m_\nu g^z$. Since z is random, E' will be a random element of $\mathcal{G}_2$ from the adversaries view and the message contains no information about m_ν.

Phase 2 The simulator acts exactly as it did in Phase 1.

Guess $\mathcal{A}$ will submit a guess ν' of ν. If $\nu = \nu'$ the simulator will output $\mu' = 0$ to indicate that it was given a Modified BLDDH-tuple otherwise it will output $\mu' = 1$ to indicate it was given a random 4-tuple.

As shown in the construction the simulator's generation of public parameters and private keys is identical to that of the actual scheme.

In the case where $\mu = 1$ the adversary gains no information about ν. Therefore, we have $\Pr[\nu \neq \nu' | \mu = 1] = \frac{1}{2}$. Since the simulator guesses $\mu' = 1$ when $\nu \neq \nu'$, we have $\Pr[\mu' = \mu | \mu = 1] = \frac{1}{2}$.

If $\mu = 0$ then the adversary sees an encryption of m_ν. The adversary's advantage in this situation is ϵ by definition. Therefore, we have $\Pr[\nu = \nu' | \mu = 0] = \frac{1}{2} + \epsilon$. Since the simulator guesses $\mu' = 0$ when $\nu = \nu'$, we have $\Pr[\mu' = \mu | \mu = 0] = \frac{1}{2} + \epsilon$.

The overall advantage of the simulator in the Modified BLDDH game is $\frac{1}{2}\Pr[\mu' = \mu | \mu = 0] + \frac{1}{2}\Pr[\mu' = \mu | \mu = 1] - \frac{1}{2} = \frac{1}{2}(\frac{1}{2} + \epsilon) + \frac{1}{2}\frac{1}{2} - \frac{1}{2} = \frac{1}{2}\epsilon$. □

Theorem 1. *Our scheme is Secure in the Multiple Fuzzy Selective-ID model.*

Proof. The proof follows immediately from Lemmas 2 and 1. □

5.1 Chosen-Ciphertext Security

Our security definitions and proofs have been in the chosen-plaintext model. Our scheme can be extended to the chosen-ciphertext model by applying the technique of using simulation-sound NIZK proofs to achieve chosen-ciphertext security [15, 16, 11] as described by Canetti et al. [3].

5.2 Security in Standard IBE Model

While our main security proof is in the slightly weaker Selective-ID model we conjecture that our scheme is secure in the standard IBE model of security. In this model the adversary commits to the identity that he will attack after seeing the public parameters and receiving chosen private keys. We assert that our scheme can be reduced to a much stronger (non-standard) assumption in which the adversary makes interactive queries to an oracle.

6 Conclusion

We introduced the concept of Fuzzy Identity Based Encryption, which allows for error-tolerance between the identity of a private key and the public key used to encrypt a ciphertext. Fuzzy Identity Based Encryption has a direct application to Identity Based Encryption using a biometric as a public key.

We presented a construction of a Fuzzy IBE scheme that uses Hamming distance as the distance metric between identities. Finally, we proved our scheme under the Selective-ID model by reducing it to an assumption that can be viewed as a modified version of the Bilinear Decisional Diffie-Hellman assumption.

References

[1] Dan Boneh and Xavier Boyen. Efficient selective-id secure identity based encryption without random oracles. In *Proceedings of the International Conference on Advances in Cryptology (EUROCRYPT '04)*, Lecture Notes in Computer Science. Springer Verlag, 2004.

[2] Dan Boneh and Matthew K. Franklin. Identity-based encryption from the weil pairing. In *Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology*, pages 213–229. Springer-Verlag, 2001.

[3] Ran Canetti, Shai Halevi, and Jonathan Katz. A forward-secure public-key encryption scheme. In *Proceedings of Eurocrypt 2003*. Springer-Verlag, 2003.

[4] G.I. Davida, Y. Frankel, and B.J. Matt. On enabling secure applications through off-line biometric identification. *In IEEE Symposium on Privacy and Security*, 1998.

[5] Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate string keys from biometrics and other noisy data. In *Proceedings of the International Conference on Advances in Cryptology (EUROCRYPT '04)*, Lecture Notes in Computer Science. Springer Verlag, 2004.

[6] Craig Gentry and Alice Silverberg. Hierarchical id-based cryptography. In *Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security*, pages 548–566. Springer-Verlag, 2002.

[7] Philippe Golle and Jessica Staddon. Personal communication.

[8] Jeremy Horwitz and Ben Lynn. Toward hierarchical identity-based encryption. In *Advances in Cryptology: EUROCRYPT 2002*, pages 466–481, 2002.

[9] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Proceedings of the 6th ACM conference on Computer and communications security*, pages 28–36. ACM Press, 1999.

[10] Ari Juels and Martin Wattenberg. A fuzzy vault scheme. In *Proceedings of IEEE International Symposium on Information Theory*, 2002.

[11] Yehuda Lindell. A simpler cosntruction of CCA2-secure public-key encryption under general asumptions. In *In Proceedings of Eurocrypt 2003*, 2003.

[12] Fabian Monrose, Michael K. Reiter, Q. (Peter) Li, Daniel Lopresti, and Chilin Shih. Towards voice generated cryptographic keys on resource constrained devices. In *Proceedings of the 11th USENIX Security Symposium*, 2002.

[13] Fabian Monrose, Michael K. Reiter, Q. (Peter) Li, and Susanne Wetzel. Cryptographic key generation from voice. In *Proceedings of the IEEE Conference on Security and Privacy*, 2001.

[14] Fabian Monrose, Michael K. Reiter, and Susanne Wetzel. Password hardening based on keystroke dynamics. In *Proceedings of the 6th ACM conference on Computer and communications security*, pages 73–82. ACM Press, 1999.

[15] Moni Naor and Moti Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *ACM Symposium on Theory of Computing*, pages 427–437, 1990.

[16] Amit Sahai. Non-malleable non-interactive zero knowledge and adaptive chosen-ciphertext security. In *In Proceedings of 40 IEEE Symp. on Foundations of Computer Science*, 1999.

[17] Adi Shamir. How to share a secret. *Communications. ACM*, 22(11):612–613, 1979.

[18] Adi Shamir. Identity-based cryptosystems and signature schemes. In *Proceedings of CRYPTO 84 on Advances in cryptology*, pages 47–53. Springer-Verlag New York, Inc., 1985.

A Security Game Reduction

Lemma 2. *If a Fuzzy IBE scheme is secure according to the Multiple Fuzzy Selective-ID model of security then it is also secure in the Basic Fuzzy Selective-ID model of security.*

Proof. Suppose we have an adversary $\mathcal{A}$ that is successful with advantage ϵ in the Multiple Fuzzy Selective-ID game for a certain scheme. Furthermore, suppose $\mathcal{A}$ picks at most l target identities in the first phase. We can then build a simulator $\mathcal{B}$ that is successful with advantage at least $\frac{\epsilon}{2l}$ in the Basic Fuzzy Selective-ID game.

We first define hybrid experiments for the Multiple Fuzzy Selective-ID game.

For $0 \leq i \leq l$ we define H_i as a game in which the adversary $\mathcal{A}$ is given an encryption of a random message for the first i ciphertexts (encrypted under the respective first i identities) and the rest of the ciphertexts are determined by the coin flip b as in the defined game. $\mathcal{A}$'s advantage in H_0 is ϵ by our assumption and $\mathcal{A}$'s advantage in H_l is 0 since the ciphertexts will be completely independent of b from the adversary's point of view. Therefore, by the triangle inequality there exists a $0 \leq j < l$ such that $\mathcal{A}$ has a greater advantage in H_j than H_{j+1} by at least $\frac{\epsilon}{l}$.

We construct an adversary $\mathcal{B}$ that plays the Basic Fuzzy Selective-ID game in the following way:

Setup $\mathcal{B}$ will first run $\mathcal{A}$ and get $l' \leq l$ target identities $\alpha_1, \ldots, \alpha_{l'}$. $\mathcal{B}$ will then submit α_{j+1} as the target identity for its game. $\mathcal{B}$ then receives the public parameters for its scheme and passes these on to $\mathcal{A}$.

Phase 1 $\mathcal{B}$ responds to private key queries of $\mathcal{A}$ by passing the same private key queries to the challenger and passing the results back to $\mathcal{A}$. The only private key queries that $\mathcal{B}$ is not allowed to make are within d Hamming distance of α_{j+1}. These are also not allowed by $\mathcal{A}$. Therefore, $\mathcal{B}$ can respond to all legal queries made by $\mathcal{A}$.

Challenge In the challenge phase $\mathcal{A}$ will submit $m_{1,0}, \ldots, m_{l',0}$ and $m_{1,1}, \ldots, m_{l',1}$. $\mathcal{B}$ will select a random message, R, and flip a coin b. $\mathcal{B}$ then submits $m_0 = R$ and $m_1 = m_{j+1,b}$ as the ciphertexts it wishes to be challenged on in its game. The challenger will flip a coin β outside $\mathcal{B}$'s view. $\mathcal{B}$ will then receive back the ciphertext C which is an encryption of m_β with the key α_{j+1} from the challenger.

$\mathcal{B}$ then assigns ciphertexts $c_1, \ldots, c_{l'}$ int the following way. It creates $c_1, \ldots, c_j$ as the message R encrypted under the respective identities $\alpha_1, \ldots, \alpha_j$. It assigns $c_{j+1} = C$. Finally, it creates $c_{j+2}, \ldots, c_{l'}$ by encrypting $m_{j+2,b}, \ldots, m_{l',b}$ under $\alpha_{j+2}, \ldots, \alpha_{l'}$ respectively. (If $j \geq l'$ then only the first step needs to be taken.) The ciphertexts are then passed to $\mathcal{A}$.

Phase 2 Subsequent private key queries are then satisfied as in Phase 1.

Guess $\mathcal{A}$ then outputs its guess b'. If $b = b'$ then $\mathcal{B}$ outputs its guess $\beta' = 1$, otherwise it outputs $\beta' = 0$.

If $\beta = 1$ then C is an encryption of $m_{j+1,b}$ and $\mathcal{A}$ will be playing the hybrid game H_j. Otherwise, if $\beta = 0$ then C is an encryption of R and $\mathcal{A}$ is playing the hybrid game H_{j+1}. It follows that $\Pr[b' = b|\beta = 1] - \Pr[b' = b|\beta = 0] \geq \frac{\epsilon}{l}$. Since $\beta' = 1$ iff $b = b'$, we have $\Pr[\beta' = 1|\beta = 1] - \Pr[\beta' = 1|\beta = 0] \geq \frac{\epsilon}{l}$. Therefore, $\mathcal{B}$ has an advantage of at least $\frac{\epsilon}{2l}$ in the Basic Fuzzy Selective-ID game. □