Fuzzy Identity-Based Encryption

Amit Sahai[1]* and Brent Waters[2]

[1] University of California, Los Angeles
sahai@cs.ucla.edu
[2] Stanford University
bwaters@cs.stanford.edu

Abstract. We introduce a new type of Identity-Based Encryption (IBE) scheme that we call Fuzzy Identity-Based Encryption. In Fuzzy IBE we view an identity as set of descriptive attributes. A Fuzzy IBE scheme allows for a private key for an identity, ω, to decrypt a ciphertext encrypted with an identity, ω', if and only if the identities ω and ω' are close to each other as measured by the "set overlap" distance metric. A Fuzzy IBE scheme can be applied to enable encryption using biometric inputs as identities; the error-tolerance property of a Fuzzy IBE scheme is precisely what allows for the use of biometric identities, which inherently will have some noise each time they are sampled. Additionally, we show that Fuzzy-IBE can be used for a type of application that we term "attribute-based encryption".
In this paper we present two constructions of Fuzzy IBE schemes. Our constructions can be viewed as an Identity-Based Encryption of a message under several attributes that compose a (fuzzy) identity. Our IBE schemes are both error-tolerant and secure against collusion attacks. Additionally, our basic construction does not use random oracles. We prove the security of our schemes under the Selective-ID security model.

1 Introduction

Identity-Based Encryption [15] (IBE) allows for a sender to encrypt a message to an identity without access to a public key certificate. The ability to do public key encryption without certificates has many practical applications. For example, a user can send an encrypted mail to a recipient, e.g. bobsmith@gmail.com, without the requiring either the existence of a Public-Key Infrastructure or that the recipient be on-line at the time of creation.

One common feature of all previous Identity-Based Encryption systems is that they view identities as a string of characters. In this paper we propose a new type of Identity-Based Encryption that we call *Fuzzy Identity-Based Encryption* in which we view identities as a set of descriptive attributes. In a Fuzzy Identity-Based Encryption scheme, a user with the secret key for the identity ω is able to decrypt a ciphertext encrypted with the public key ω' if and only if ω and ω'

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

are within a certain distance of each other as judged by some metric. Therefore, our system allows for a certain amount of error-tolerance in the identities.

Fuzzy-IBE gives rise to two interesting new applications. The first is an Identity-Based Encryption system that uses biometric identities. That is we can view a user's biometric, for example an iris scan, as that user's identity described by several attributes and then encrypt to the user using their biometric identity. Since biometric measurements are noisy, we cannot use existing IBE systems. However, the error-tolerance property of Fuzzy-IBE allows for a private key (derived from a measurement of a biometric) to decrypt a ciphertext encrypted with a slightly different measurement of the same biometric.

Secondly, Fuzzy IBE can be used for an application that we call "attribute-based encryption". In this application a party will wish to encrypt a document to all users that have a certain set of attributes. For example, in a computer science department, the chairperson might want to encrypt a document to all of its systems faculty on a hiring committee. In this case it would encrypt to the identity {"hiring-committee","faculty","systems"}. Any user who has an identity that contains all of these attributes could decrypt the document. The advantage to using Fuzzy IBE is that the document can be stored on an simple untrusted storage server instead of relying on trusted server to perform authentication checks before delivering a document.

We further discuss the usefulness of using biometrics in Identity-Based and then discuss our contributions.

Using biometrics in Identity-Based Encryption In many situations, using biometric-based identity in an IBE system has a number of important advantages over "standard" 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 standard 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 is not if it is used as a public key. We must only guarantee that an attacker cannot fool the key authority into believing that an attacker owns a biometric identity that he does not.

Also, a biometric identity is an inherent trait and will always with a person. Using biometrics in Identity-Based Encryption will mean that the person will always have their public key handy. In several situations a user will want to present an encryption key to someone when they are physically present. For example, consider the case when a user is traveling and another party encrypts an ad-hoc meeting between them.

Finally, 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 standard identities, such as the name "Bob Smith" will clearly not be unique or change owners over time.

Security Against Collusion Attacks In addition to providing error-tolerance in the set of attributes composing the identity any IBE scheme that encrypts to multiple attributes must provide security against collusion attacks. In particular, no group of users should be able to combine their keys in such a way that they can decrypt a ciphertext that none of them alone could. This property is important for security in both biometric applications and "attribute-based encryption".

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.

Our primary technique is that we construct a user's private key as a set of private key components, one for each attribute in the user's identity. We share use Shamir's method of secret sharing [14] to distribute shares of a master secret in the exponents of the user's private key components. Shamir's secret sharing within the exponent gives our scheme the crucial property of being error-tolerant since only a subset of the private key components are needed to decrypt a message. Additionally, our scheme is resistant to collusion attacks. Different users have their private key components generated with different random polynomials. If multiple users collude they will be unable to combine their private key components in any useful way.

In the first version of our scheme, the public key size grows linearly with the number of potential attributes in the universe. The public parameter growth is manageable for a biometric system where all the possible attributes are defined at the system creation time. However, this becomes a limitation in a more general system where we might like an attribute to be defined by an arbitrary string. To accommodate these more general requirements we additionally provide a Fuzzy-IBE system for large universes, where attributes are defined by arbitrary strings.

We prove our scheme secure under an adapted version of the Selective-ID security model first proposed by Canetti et al. [5]. Additionally, our construction does not use random oracles. We reduce the security of our scheme to an assumption that is similar to the Decisional Bilinear Diffie-Hellman assumption.

1.1 Related Work

Identity-Based Encryption Shamir [15] first proposed the concept of Identity-Based Encryption. However, it wasn't until much later that Boneh and Franklin [3] presented the first Identity-Based Encryption scheme that was both practical and secure. Their solution made novel use of groups for which there was an efficiently computable bilinear map.

Canetti et al. [5] 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. Boneh and Boyen [2] give two schemes with improved efficiency and prove security in the Selective-ID model without random oracles.

Biometrics Other work in applying biometrics to cryptography has focused on the derivation of a secret from a biometric [12, 11, 10, 6, 9, 7, 4]. This secret can be then used for operations such as symmetric encryption or UNIX style password authentication.

The distinguishing feature of our work from the above related work on biometrics above is that we view the biometric input as potentially *public* information instead of a secret. Our only physical requirement is that the biometric cannot be imitated such that a trained human operator would be fooled. We stress the importance of this, since it is much easier to capture a digital reading of someone's biometric, than to fool someone into believing that someone else's biometric is one's own. Simply capturing a digital reading of someone's biometric would (forever) invalidate approaches where symmetric keys are systematically derived from biometric readings.

Attribute-based encryption Yao et al. [17] show how an IBE system that encrypts to multiple hierarchical-identities in a collusion-resistant manner implies a forward secure Hierarchical IBE scheme. They also note how their techniques for resisting collusion attacks are useful in attribute-based encryption. However, the cost of their scheme in terms of computation, private key size, and ciphertext size increases exponentially with the number of attributes.

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. Then, we describe our security assumptions. In Section 3 we show why two naive approaches do not work. We follow with a description of our construction in Section 4 and in Section 5 we prove the security of our scheme. We describe our second construction in Section 6. Finally, we conclude in Section 7.

2 Preliminaries

We begin by presenting our definition of security. We follow with a brief review of bilinear maps, and then state the complexity assumptions we use for our proofs of security.

2.1 Definitions

In this section we define our Selective-ID models of security for Fuzzy Identity Based Encryption. The 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.

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 many identities, γ_j, where $|\gamma_j \cap \alpha| < d$ for all j.

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

2.2 Bilinear Maps

We briefly review the facts about groups with efficiently computable bilinear maps. We refer the reader to previous literature [3] for more details.

Let $\mathbb{G}_1, \mathbb{G}_2$ be groups of prime order p, and let g be a generator of $\mathbb{G}_1$. We say $\mathbb{G}_1$ has an admissible bilinear map, $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$, into $\mathbb{G}_2$ if the following two conditions hold. The map is bilinear; for all a, b we have $e(g^a, g^b) = e(g, g)^{ab}$. The map is non-degenerate; we must have that $e(g, g) \neq 1$.

2.3 Complexity Assumptions

We state our complexity assumptions below.

Definition 2 (Decisional Bilinear Diffie-Hellman (BDH) Assumption). *Suppose a challenger chooses $a, b, c, z \in \mathbb{Z}_p$ at random. The Decisional BDH assumption is that no polynomial-time 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.*

Definition 3 (Decisional Modified Bilinear Diffie-Hellman (MBDH) Assumption). *Suppose a challenger chooses $a, b, c, z \in \mathbb{Z}_p$ at random. The Decisional MBDH assumption is that no polynomial-time adversary is to be able to distinguish the tuple $(A = g^a, B = g^b, C = g^c, Z = e(g,g)^{\frac{ab}{c}})$ from $(A = g^a, B = g^b, C = g^c, Z = e(g,g)^z)$ with more than a negligible advantage.*

3 Other Approaches

Before describing our scheme we first show three potential approaches to building a Fuzzy Identity-Based Encryption scheme and show why they fall short. This discussion additionally motivates our approach to the problem.

Correcting the error We consider the feasibility of "correcting" the errors of a biometric measurement and then use standard Identity-Based Encryption to encrypt a message under the corrected input. However, this approach relies upon the faulty assumption that each biometric input measurement is slightly deviated from some "true" value and that the set of possible "true" values are well known. In practice, the only reasonable assumption is that two measurements sampled from the same person will be within a certain distance of each other. This intuition is captured by previous work. Dodis, Rezyin, and Smith [7] use what they call a *fuzzy sketch* that contains information of a first sampling of a biometric which allows subsequent measurements to be corrected to it. If the correction could be done without any additional information then we could simply do away with the fuzzy sketch.

Key per Attribute The second naive approach we consider is for an authority to give a user a different private key for each of the attributes that describe the user. Such a system easily falls prey to simple collusion attacks where multiple users combine their keys to form identities that are a combination of their attributes. The colluders are then able to decrypt ciphertexts that none of them individually were able to decrypt.

Several Keys Suppose a key authority measures an input ω for a particular party. The authority could create a separate standard IBE private key for every ω' such that $|\omega \cap \omega'| \geq d$, for some error-tolerance parameter d. However, the private key storage will grow exponentially in d and the system will be impractical for even modest values of d.

4 Our Construction

Recall that we view identities as sets of attributes and we let the value d represent the error-tolerance in terms of minimal set overlap. When an authority is creating a private key for a user he will associate a random $d-1$ degree polynomial, $q(x)$, with each user with the restriction that each polynomial have the same valuation at point 0, that is $q(0) = y$.

For each of the attributes associated with a user's identity the key generation algorithm will issue a private key component that is tied to the user's random polynomial $q(x)$. If the user is able to "match" at least d components of the ciphertext with their private key components, then they will be able to perform decryption. However, since the private key components are tied to random polynomials, multiple user's are unable to combine them in anyway that allows for collusion attacks.

A detailed description of our scheme follows.

4.1 Description

Recall that we wish to create an IBE scheme in which a ciphertext created using identity ω can be decrypted only by a secret key ω' where $|\omega \cap \omega'| \geq d$.

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

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

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

Identities will be element subsets of some universe, $\mathcal{U}$, of size $|\mathcal{U}|$. We will associate each element with a unique integer in ${\mathbb{Z}_p}^*$. (In practice an attribute will be associated with each element so that identities will have some semantics.) Our construction follows:

Setup(d) First, define the universe, $\mathcal{U}$ of elements. For simplicity, we can take the first $|\mathcal{U}|$ elements of ${\mathbb{Z}_p}^*$ to be the universe. Namely, the integers $1, \ldots, |\mathcal{U}|$ (mod p).

Next, choose $t_1, \ldots, t_{|\mathcal{U}|}$ uniformly at random from $\mathbb{Z}_p$. Finally, choose y uniformly at random in $\mathbb{Z}_p$. The published public parameters are:

$$T_1 = g^{t_1}, \ldots, T_{|\mathcal{U}|} = g^{t_{|\mathcal{U}|}}, Y = e(g,g)^y.$$

The master key is:

$$t_1, \ldots, t_{|\mathcal{U}|}, y.$$

Key Generation To generate a private key for identity $\omega \subseteq \mathcal{U}$ the following steps are taken. A $d-1$ degree polynomial q is randomly chosen such that $q(0) = y$. The private key consists of components, $(D_i)_{i \in \omega}$, where $D_i = g^{\frac{q(i)}{t_i}}$ for every $i \in \omega$.

Encryption Encryption with the public key ω' and message $M \in \mathbb{G}_2$ proceeds as follows.

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

$$E = (\omega', E' = MY^s, \{E_i = T_i^s\}_{i \in \omega'}).$$

Note that the identity, ω', is included in the ciphertext.

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

Then, the ciphertext can be decrypted as:

$$\begin{aligned}
&E' / \prod_{i \in S} (e(D_i, E_i))^{\Delta_{i,S}(0)} \\
&= Me(g,g)^{sy} / \prod_{i \in S} \left(e(g^{\frac{q(i)}{t_i}}, g^{st_i}) \right)^{\Delta_{i,S}(0)} \\
&= Me(g,g)^{sy} / \prod_{i \in S} \left(e(g,g)^{sq(i)} \right)^{\Delta_{i,S}(0)} \\
&= M.
\end{aligned}$$

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 Efficiency and Key Sizes

The number of exponentiations in the group $\mathbb{G}_1$ to encrypt to an identity will be linear in the number of elements in the identity's description. The cost of decryption will be dominated by d bilinear map computations.

The number of group elements in the public parameters grows linearly with the number attributes in the system (elements in the defined universe). The number of group elements that compose a user's private key grow linearly with the number of attributes associated with her identity. Finally, the number of group elements in a ciphertext grows linearly with the size of the identity we are encrypting to.

4.3 Flexible Error-Tolerance

In this construction the error-tolerance is set to a fixed 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 the example of attribute-based encryption we would like to have flexibility in the number of attributes required to access a document.

There are two simple methods for achieving flexible error-tolerance. First, we can create multiple systems with different values of d and the party encrypting a message can choose the appropriate one. For m different systems the size of the public parameters and private keys both increase by a factor of m. In the second method the authority will reserve some attributes that it will issue to every key-holder as part of their identity. The party encrypting the message can increase the error-tolerance by increasing the number of these "default" attributes it includes in the encryption identity. In this approach ciphertexts must be at least as long as the maximum number of attributes that can be required in an encryption. Additionally, we can combine the above two techniques and explore tradeoffs between ciphertext size and public parameter and private key size.

5 Proof of Security

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

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

Proof. Suppose there exists a polynomial-time 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 MBDH game with advantage $\frac{\epsilon}{2}$. The simulation proceeds as follows:

We first let the challenger set the groups $\mathbb{G}_1$ and $\mathbb{G}_2$ 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 $(A, B, C, Z) = (g^a, g^b, g^c, e(g,g)^{\frac{ab}{c}})$; otherwise it sets $(A, B, C, Z) = (g^a, g^b, g^c, e(g,g)^z)$ for random a, b, c, z. We assume the universe, $\mathcal{U}$ is defined.

Init The simulator $\mathcal{B}$ runs $\mathcal{A}$ and receives the challenge identity, α.

Setup The simulator assigns the public key parameters as follows. It sets the parameter $Y = e(g, A) = e(g, g)^a$. For all $i \in \alpha$ it chooses random $\beta_i \in \mathbb{Z}_p$ and sets $T_i = C^{\beta_i} = g^{c\beta_i}$. For all $i \in \mathcal{U} - \alpha$ it chooses random $w_i \in \mathbb{Z}_p$ and sets $T_i = g^{w_i}$.

It then gives the public parameters to $\mathcal{A}$. Notice that from the view $\mathcal{A}$ all parameters are chosen at random as in the construction.

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

Suppose $\mathcal{A}$ requests a private key γ where $|\gamma \cap \alpha| < d$. We first define three sets Γ, Γ', S in the following manner:

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

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

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

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

$$\text{If } i \in \Gamma: \; D_i = g^{s_i} \text{ where } s_i \text{ is chosen randomly in } \mathbb{Z}_p.$$

$$\text{If } i \in \Gamma' - \Gamma: \; D_i = g^{\frac{\lambda_i}{w_i}} \text{ where } \lambda_i \text{ is chosen randomly in } \mathbb{Z}_p.$$

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 $q(0) = a$. For $i \in \Gamma$ we have $q(i) = c\beta_i s_i$ and for $i \in \Gamma' - \Gamma$ we have $q(i) = \lambda_i$.

The simulator can calculate the other D_i values where $i \notin \Gamma'$ since the simulator knows the discrete log of T_i for all $i \notin \alpha$. The simulator makes the assignments as follows:

$$\text{If } i \notin \Gamma': \; D_i = (\prod_{j \in \Gamma} C^{\frac{\beta_j s_j \Delta_{j,S}(i)}{w_i}})(\prod_{j \in \Gamma' - \Gamma} g^{\frac{\lambda_j \Delta_{j,S}(i)}{w_i}})Y^{\frac{\Delta_{0,S}(i)}{w_i}}$$

Using interpolation the simulator is able to calculate $D_i = g^{\frac{q(i)}{t_i}}$ for $i \notin \Gamma'$ where $q(x)$ was implicitly defined by the random assignment of the other $d-1$ variables $D_i \in \Gamma'$ and the variable Y.

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:

$$E = (\alpha, E' = M_\nu Z, \{E_i = B^{\beta_i}\}_{i \in \alpha}).$$

If $\mu = 0$, then $Z = e(g,g)^{\frac{ab}{c}}$. If we let $r' = \frac{b}{c}$, then we have $E_0 = M_\nu Z = M_\nu e(g,g)^{\frac{ab}{c}} = M_\nu e(g,g)^{ar'} = M_\nu Y^{r'}$ and $E_i = B^{\beta_i} = g^{b\beta_i} = g^{\frac{b}{c}c\beta_i} = g^{r'c\beta_i} = (T_i)^{r'}$. Therefore, the ciphertext is a random encryption of the message m_ν under the public key α.

Otherwise, if $\mu = 1$, then $Z = g^z$. We then have $E' = M_\nu e(g,g)^z$. Since z is random, E' will be a random element of $\mathbb{G}_2$ 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 MBDH-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 Decisional MBDH game is $\frac{1}{2}\Pr[\mu' = \mu | \mu = 0] + \frac{1}{2}\Pr[\mu' = \mu | \mu = 1] - \frac{1}{2} = \frac{1}{2}(\frac{1}{2} + \epsilon) + \frac{1}{2}\frac{1}{2} - \frac{1}{2} = \frac{1}{2}\epsilon$. □

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 [13]. Alternatively, if we are willing to use random oracles, then the we can use standard techniques such as the Fujisaki-Okamoto transformation [8].

5.2 Security in Full IBE Model

Suppose all identities are composed of n attributes and we have a universe of attributes, $\mathcal{U}$. We make the observation [2] that our scheme is secure in the full model with a factor of $\binom{|\mathcal{U}|}{n}$ in the reduction.

The original IBE scheme of Boneh and Franklin [3] and a later schemes of Boneh and Boyen [2] and Waters [16] achieve IBE in the full model with non-exponential reductions. However, all methods achieve this by essentially removing the relationships between nearby identities. In Fuzzy-IBE it is essential that there exists a relationship between nearby identities. Therefore, we conjecture that a scheme that has a non-exponential loss of security in the full model will require significantly different methods than those seen in prior work.

6 Large Universe Construction

In the previous construction the size of the public parameters grows linearly with the number of possible attributes in the universe. We describe a second scheme which uses all elements of $\mathbb{Z}_p{}^*$ as the universe, yet the public parameters only grow linearly in a parameter n, which we fix as the maximum size identity we can encrypt to.

In addition to decreasing the public parameter size, having a large universe allows us to apply a collision-resistant hash function $H : \{0,1\}^* \rightarrow \mathbb{Z}_p{}^*$ and use arbitrary strings as attributes. We can now use attributes that were not necessarily considered during the public key setup. For example, we can add any verifiable attribute, such as "Ran in N.Y. Marathon 2005", to a user's private key.

Our large universe construction is built using similar concepts to the previous scheme and uses an algebraic technique of Boneh and Boyen [2]. Additionally, we reduce the security of this scheme to the Decisional BDH problem. We now describe our construction and give our proof of security.

6.1 Description

Let $\mathbb{G}_1$ be bilinear group of prime order p, and let g be a generator of $\mathbb{G}_1$. Additionally, let $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ denote the bilinear map. We restrict encryption identities to be of length n for some fixed n.

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

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

Identities will be sets of n elements of $\mathbb{Z}_p{}^*$.[3] 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 $\mathbb{Z}_p{}^*$. Our construction follows:

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

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

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

We can view T as the function $g_2^{x^n} g^{h(x)}$ for some n degree polynomial h. The public key is published as: $g_1, g_2, t_1, \ldots, t_{n+1}$ and the private key is y.

Key Generation To generate a private key for identity ω 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, $\{D_i\}_{i \in \omega}$, where the elements are constructed as

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

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

The other set is $\{d_i\}_{i \in \omega}$ where the elements are constructed as

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

[3] With some minor modifications to our scheme, which we omit for simplicity, we can encrypt to all identities of size $\leq n$.

Encryption Encryption with the public key ω' and message $M \in \mathbb{G}_2$ proceeds as follows.

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

$$E = (\omega', E' = Me(g_1, g_2)^s, E'' = g^s, \{E_i = T(i)^s\}_{i \in \omega'}).$$

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

Then, the ciphertext can be decrypted as:

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

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.

6.2 Efficiency and Key Sizes

Again, he number of exponentiations in the group $\mathbb{G}_1$ to encrypt to an identity will be linear in the number of elements in the identity's description. The cost of decryption will be dominated by $2 \cdot d$ bilinear map computations.

The key feature of the scheme is that the number of group elements in the public parameters only grows linearly with, n, the maximum number of attributes that can describe an encryption identity. The number of group elements that compose a user's private key grow linearly with the number of attributes associated with her identity. Finally, the number of group elements in a ciphertext grows linearly with the size of the identity we are encrypting to.

6.3 Proof of Security

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

Theorem 2. *If an adversary can break our scheme in the Fuzzy Selective ID Model, then a simulator can be constructed to play the Decisional BDH game with a non-negligible advantage.*

Proof. Suppose there exists a polynomial-time 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 BDH game with advantage $\frac{\epsilon}{2}$.

The simulation proceeds as follows:

We first let the challenger set the groups $\mathbb{G}_1$ and $\mathbb{G}_2$ 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.

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

Setup The simulator assigns the public parameters $g_1 = A$ and $g_2 = B$. It then chooses a random n degree polynomial $f(x)$ and calculates an n degree polynomial $u(x)$ such that $u(x) = -x^n$ for all $x \in \alpha$ and where $u(x) \neq -x^n$ for some other x. Since $-x^n$ and $u(x)$ are two n degree polynomials they will either agree on at most n points or they are the same polynomial. Our construction assures that $\forall x\ u(x) = -x^n$ if and only if $x \in \alpha$.

Then, for i from 1 to $n+1$ the simulator sets $t_i = g_2^{u(i)} g^{f(i)}$. Note that since $f(x)$ is a random n degree polynomial all t_i will be chosen independently at random as in the construction and we implicitly have $T(x) = g_2^{i^n + u(i)} g^{f(i)}$.

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 $\mathcal{A}$ requests a private key γ. We first define three sets Γ, Γ', S 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:

$$D_i = g_2^{\lambda_i} T(i)^{r_i} \text{ where } r_i, \lambda_i \text{ are chosen randomly in } \mathbb{Z}_p \text{ and we let } d_i = g^{r_i}.$$

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 in Γ randomly by setting $q(i) = \lambda_i$ in addition to having $q(0) = a$.

The simulator also needs to calculate the decryption key values for all $i \in \gamma - \Gamma'$. We calculate these points to be consistent with our implicit choice of $q(x)$. The key components are calculated as:

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

and

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

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 our construction of $u(x)$.

Let $r_i = (r_i' - \frac{a}{i^n+u(i)})\Delta_{0,S}(i)$ and let $q(x)$ be defined as above. We then have:

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

Additionally, we have:

$$d_i = (g_1^{\frac{-1}{i^n+u(i)}} g^{r_i'})^{\Delta_{0,S}(i)} = (g^{r_i' - \frac{a}{i^n+u(i)}})^{\Delta_{0,S}(i)} = g^{r_i}$$

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 since our choices of λ_i induce a random $d-1$ degree polynomial and our construction of the private keys components d_i and D_i.

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:

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

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

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

This is a valid ciphertext for the message M_ν under the identity α.

Otherwise, if $\mu = 1$, then $Z = e(g,g)^z$ and $E' = M_\nu e(g,g)^z$. Since z is random, E' will be a random element of $\mathbb{G}_2$ 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 BDH-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 Decisional BDH game is $\frac{1}{2}\Pr[\mu' = \mu | \mu = 0] + \frac{1}{2}\Pr[\mu' = \mu | \mu = 1] - \frac{1}{2} = \frac{1}{2}(\frac{1}{2} + \epsilon) + \frac{1}{2}\frac{1}{2} - \frac{1}{2} = \frac{1}{2}\epsilon$. □

7 Conclusions

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. We described two practical applications of Fuzzy-IBE of encryption using biometrics and attribute-based encryption.

We presented our construction of a Fuzzy IBE scheme that uses set overlap 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.

This work motivates a few interesting open problems. The first is whether it is possible to create a Fuzzy IBE scheme where the attributes come from multiple authorities. While, it is natural for one authority to certify all attributes that compromise a biometric, in attribute-based encryption systems there will often not be one party that can act as an authority for all attributes. Also, a Fuzzy-IBE scheme that hides the public key that was used to encrypt the ciphertext [1] is intriguing. Our scheme uses set-overlap as a similarity measure between identities. (We note a Hamming-distance construction can also be built using our techniques.) An open problem is to build other Fuzzy-IBE schemes that use different distance metrics between identities.

Acknowledgments

We would like to thank Don Coppersmith, Ed Felten and Philippe Golle, Ari Juels, and the reviewers of Eurocrypt 2005 for providing helpful comments and suggestions.

References

1. Mihir Bellare, Alexandra Boldyreva, Anand Desai, and D. Pointcheval. Key-privacy in public-key encryption. *Lecture Notes in Computer Science*, 2248, 2001.

2. 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.
3. 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.
4. Xavier Boyen. Reusable cryptographic fuzzy extractors. In *ACM Conference on Computer and Communications Security—CCS 2004*, 2004.
5. Ran Canetti, Shai Halevi, and Jonathan Katz. A forward-secure public-key encryption scheme. In *Proceedings of Eurocrypt 2003*. Springer-Verlag, 2003.
6. 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.
7. 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.
8. Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pages 537–554. Springer-Verlag, 1999.
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. 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.
11. 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.
12. 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.
13. 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.
14. Adi Shamir. How to share a secret. *Communications. ACM*, 22(11):612–613, 1979.
15. 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.
16. Brent Waters. Efficient identity based encryption without random oracles. In *To Appear in Proceedings Eurocrypt 2005*, 2005.
17. Danfeng Yao, Nelly Fazio, Yevgeniy Dodis, and Anna Lysyanskaya. Id-based encryption for complex hierarchies with applications to forward security and broadcast encryption. In *ACM Conference on Computer and Communications Security—CCS 2004*, 2004.