Definitional Issues in Functional Encryption

Adam O'Neill*

Abstract

We provide a formalization of the emergent notion of "functional encryption," as well as introduce various security notions for it, and study relations among the latter. In particular, we show that indistinguishability and semantic security based notions of security are *inequivalent* for functional encryption in general; in fact, "adaptive" indistinguishability does not even imply "non-adaptive" semantic security. This is alarming given the large body of work employing (special cases of) the former. We go on to show, however, that in the "non-adaptive" case an equivalence does hold between indistinguishability and semantic security for what we call *preimage sampleable* schemes. We take this as evidence that for preimage sampleable schemes an indistinguishability based notion may be acceptable in practice. We show that some common functionalities considered in the literature satisfy this requirement.

1 Introduction

Functional encryption. In recent years, a notion of "functional encryption" (FE) has emerged as a new paradigm for public-key encryption, wherein a receiver, given a ciphertext, is able to learn certain functions of the underlying message based on its secret keys (not necessarily the decryption). Special cases of FE include identity-based encryption [BF03], public-key encryption with keyword search [BCOP04, ABC+08], attribute-based encryption [SW05, GPSW06, BSW07], and predicate encryption [BW07, KSW08, LOS+10, OT10].[1] However, a general study of FE and its security seems not to have appeared. Here we initiate one, and in doing so we uncover some interesting definitional issues that cause one to re-evaluate what exactly is being achieved by this body of work.

Syntax and security notions. First we give a syntactic definition of FE, which extends that for predicate encryption introduced by Boneh and Waters [BW07]. We then formulate an "indistinguishability based" notion of privacy (IND), which again extends the security notion for predicate encryption introduced in [BW07]. Informally, the IND notion asks that it be hard for an adversary to distinguish between the encryptions of any two messages that agree on all the functions corresponding to the secret keys it requested. We go on to introduce a more complicated but more natural "semantic security based" (SS) notion of privacy in the spirit of the classical notion for public-key encryption [GM84], to capture the intuition that anything the adversary can compute from a ciphertext it could as well compute from the evaluations of the functions corresponding to

*University of Texas at Austin. Work done in part while the author was a Ph.D. student at Georgia Institute of Technology.

[1] We do not mean here to claim credit for the general concept of functional encryption and its generalizing these primitives; indeed, this view was present in prior work (e.g., [SW05, BW07, KSW08]) and in a talk by Waters [Wat].

the secret keys it requested on underlying message. We note that a novel feature of our definitions, which turns out to be important when considering relations among them, is that they distinguish between *adaptive* and *non-adaptive* access to the secret-key derivation oracle to aid in the adversary's task (roughly, this distinction is analogous to that between access to the decryption oracle in adaptive versus non-adaptive chosen-ciphertext attack, see e.g. [BDPR98]).

Relations among security notions. In the classical setting of public-key encryption, semantic security and indistinguishability based formulations of security are well-known to be equivalent [MRS88]. We ask whether the same is true for FE. Surprisingly, we show that IND under adaptive access to the secret-key derivation oracle does not imply SS even under *non-adaptive* such access. To see why, consider a functional encryption scheme for a single function f. But suppose there is another function g that has the same "equality pattern" as f on the message space (i.e., two messages have the same f-value just when they have the same g-value). Furthermore, suppose $g(m)$ is hard to compute given $f(m)$. Now, if the functional encryption scheme is such that the secret keys created by the scheme, which are supposed to allow computing f, also allow computing g, the scheme is certainly not semantically secure. However, an IND adversary is "bound" to choosing messages that agree on f, hence also on g, and so cannot use computing g to its advantage. Our counter-example formalizes and generalizes this intuition. Another shortcoming of the IND notion we observe is that it is essentially vacuous[2] for some functions, such as a collision-resistant hash function. Then, it is hard for the adversary to find two messages that agree on the function.

Achievability. Finally, we ask the question of whether the SS notion for FE is *achievable*. In particular, we note that achieving SS under adaptive access to the key derivation oracle seems difficult. In the proof, the simulator seemingly must choose a "dummy" ciphertext on which to run the adversary *before* knowing what values the challenge message should have when evaluated under the functions for which the adversary will later request secret keys. Intuitively, this means the number of possible keys for a given function should at least be as large as the number of possible outputs of the latter. This situation is reminiscent of that for (non-interactive) non-committing encryption, for which impossiblity results are known [Nie02]. However, it is unclear to us how to formalize this connection since there could be other proof techniques.

We do, however, obtain positive results in the case of *non-adaptive* SS. Here we identify a key property of functional encryption schemes that we call *preimage sampleability*. Intuitively, this means that, given the function values of some underlying message, it should always be possible to efficiently find *some* message consistent with them. We show that for preimage sampleable FE schemes, IND is *equivalent* to SS (both under non-adaptive access to the key-derivation oracle). (Thus, for non-adaptive securiy our above-mentioned counter-example is tight.) One reason we believe this is important is that non-adaptive SS suffices to rule out the "pathological" examples of schemes we gave that meet IND but not SS.[3] We take this as evidence that, for preimage sampleable schemes, IND (under adaptive access to the key-derivation oracle) may be acceptable in practice. We conclude by showing that some common function classes considered in the literature, including the powerful inner-product predicates realized in [KSW08, LOS+10, OT10], are preimage sampleable.

[2] At least, it is vacuous with respect to attacks that require the adversary to query its key derivation oracle; i.e., attacks where the adversary actually uses the secret keys. A functional encryption scheme may of course already not be semantically secure in the classical sense.

[3] On the other hand, it is possible to extend them to even more extreme examples that violate SS only under adaptive access to the key-derivation oracle, but these start to really stretch plausibility.

CONCURRENT AND INDEPENDENT WORK. Independently of our work, Boneh et al. [BSW11] also undertake a general study of FE. In particular they give a syntactic definition as well as indistinguishability and semantic security based formulations of privacy. They also give a counter-example showing that IND does not imply SS.[4] They also formalize the connection between SS under adaptive access to the key derivation oracle (although they do not distinguish between adaptive versus non-adaptive here) and non-committing encryption, showing via an argument in the style of [Nie02] that the former is not achievable at all without (programmable) random oracles but is achievable in the random oracle model. We feel this further highlights the importance of our results on the (standard model) achievability of non-adaptive SS.

In another concurrent and independent work, Chase and Kamara [CK10] introduced and studied a notion of "structural encryption," which they observe is similar to FE except that it is in the symmetric-key setting and secret keys "work" for only one specific ciphertext on which the former is dependent. They employ a semantic security based definition of security and also note a connection to non-committing encryption, namely that under their definition secret keys should be as long as the number of possible outputs of the associated function. Note that the reason that Boneh et al. [BSW11] obtain an impossibility result for FE is that a secret key must work for *all* ciphertexts.

2 Functional Encryption and its Security

We define the syntax of functional encryption and various security notions for it.

2.1 Syntax

A *functional encryption scheme* for the class of PT functions (aka. functionality) $\mathcal{F}$ on message-space Σ (both of which implicitly depend on k) is a tuple of algorithms $\mathcal{FE} = (\mathsf{Setup}, \mathsf{KDer}, \mathsf{Enc}, \mathsf{Eval})$ such that:

- Setup on input 1^k outputs a *master public key* pk and *master secret key* sk.
- KDer on input the master secret key sk and a (description of a) function $f \in \mathcal{F}$ outputs an *evaluation token* (aka. secret key) sk_f for f.
- Enc on input a public key pk and a message (aka. attribute) $m \in \Sigma$ outputs a ciphertext c.
- Eval on input an evaluation token sk_f and a ciphertext c outputs a string y or $\perp$.

For correctness we require that for all $k \in \mathbb{N}$, all $f \in \mathcal{F}$, and all $m \in \Sigma$,

$$\mathsf{Eval}(sk_f, \mathsf{Enc}(pk, m)) = f(m)$$

with probability 1 over $(pk, sk) \xleftarrow{\$} \mathsf{Setup}(1^k)$ and $sk_f \xleftarrow{\$} \mathsf{KDer}(sk, f)$.

Note that this notion is in particular a generalization of identity-based encryption (IBE) [BF03], public-key encryption with keyword search (PEKS) [BCOP04, ABC+08], attribute-based encryption (ABE) [SW05, GPSW06, BSW07], and predicate encryption (PE) [BW07, KSW08, LOS+10, OT10]. For example, in the case of identity-based encryption, the "message" would consist of the identity concatenated with the actual payload, and the secret key would be associated with the function $f_{ID}(ID' \| x) = x$ if $ID = ID'$ and $\perp$ otherwise.

[4] Our counter-example is slightly more general. In particular, we observe a separation even for schemes (such as those proposed in the literature) where the adversary can find two messages that agree on the functions corresponding to the secret keys it requested.

2.2 Security Definitions

We present various formulations of privacy for functional encryption. Broadly, the definitions are either *indistinguishability based* or *semantic-security based*. In each case we also define a *token non-adaptive* (TNA) variant, where the adversary gets access to a token derivation oracle only before it sees the challenge ciphertext.

Regarding special cases, we note that our security notions yield the *anonymous* (aka. attribute-hiding) versions of IBE, ABE, and PE, where the identity or attribute is hidden by the ciphertext (or their "predicate-only" counterparts following the terminology of [KSW08]). The contemporaneous work of [BSW11] provides a more general and comprehensive treatment.

INDISTINGUISHABILITY BASED PRIVACY. The indistinguishability-based formulation follows [BW07] and tries to capture the intuition that the adversary is unable to distinguish between the encryptions of two different messages that it cannot trivially distinguish using its tokens. Let $\mathcal{FE} = (\mathsf{Setup}, \mathsf{KDer}, \mathsf{Enc}, \mathsf{Eval})$ be a functional encryption scheme for the class of functions $\mathcal{F}$ over message-space Σ and let $A = (A_1, A_2)$ be an adversary. For mode $\in \{\text{full}, \text{tna}\}$ [5] and $k \in \mathbb{N}$ we associate to $\mathcal{FE}$ and A the experiments

$$
\begin{array}{l}
\textbf{Experiment } \mathbf{Exp}^{\text{ind-mode}}_{\mathcal{FE},A}(k)\text{:} \\
\quad b \xleftarrow{\$} \{0,1\} \\
\quad (pk, sk) \xleftarrow{\$} \mathsf{Setup}(1^k) \\
\quad (m_0, m_1, st) \xleftarrow{\$} A_1^{\mathsf{KDer}(sk,\cdot)}(pk) \\
\quad c \xleftarrow{\$} \mathsf{Enc}(pk, m_b) \\
\quad b' \xleftarrow{\$} A_2^{\mathcal{O}(sk,\cdot)}(pk, c, st) \\
\quad \text{If } b = b' \text{ return 1 else return 0}
\end{array}
$$

where if mode = full then $\mathcal{O}(sk, \cdot) = \mathsf{KDer}(sk, \cdot)$ and if mode = tna then $\mathcal{O}(sk, \cdot) = \varepsilon$ (the empty oracle). We require that $|m_0| = |m_1|$ and every query f that A_1 or A_2 makes to its oracle satisfies $f(m_0) = f(m_1)$. Denote by $\Pr\left[\, \mathbf{Exp}^{\text{ind-mode}}_{\mathcal{FE},A}(k) = 1 \,\right]$ the probability that the corresponding IND-MODE experiment outputs 1, and define

$$
\mathbf{Adv}^{\text{ind-mode}}_{\mathcal{FE},A}(k) = 2 \cdot \Pr\left[\, \mathbf{Exp}^{\text{ind-mode}}_{\mathcal{FE},A}(k) = 1 \,\right] - 1 \; .
$$

We say that $\mathcal{FE}$ is *IND-MODE secure* if $\mathbf{Adv}^{\text{ind-mode}}_{\mathcal{FE},A}(\cdot)$ is negligible for all PPT adversaries A.

SEMANTIC-SECURITY BASED PRIVACY. The semantic-security formulation is new and tries to capture the intuition that anything the adversary can compute from a ciphertext and the tokens it can compute from the tokens and the values of the corresponding functions on the underlying message. Let $\mathcal{FE} = (\mathsf{Setup}, \mathsf{KDer}, \mathsf{Enc}, \mathsf{Eval})$ be a functional encryption scheme for the class of functions $\mathcal{F}$ over message-space Σ, let $A = (A_1, A_2, A_3)$ be an adversary, let S be a simulator. For mode $\in \{\text{full}, \text{tna}\}$ and $k \in \mathbb{N}$ we associate to $\mathcal{FE}$, A, and S the experiments

[5] We stress that our use of the terminology "full" security differs from the literature in that it refers to adaptive access to the key derivation oracle rather than adaptive choice of the challenge messages.

Experiment $\mathbf{Exp}^{\text{ss-real-mode}}_{\mathcal{FE},A}(k)$:	**Experiment** $\mathbf{Exp}^{\text{ss-ideal-mode}}_{\mathcal{FE},A,S}(k)$:
$(pk, sk) \xleftarrow{\$} \mathsf{Setup}(1^k)$	$(pk, sk) \xleftarrow{\$} \mathsf{Setup}(1^k)$
$st \xleftarrow{\$} A_1^{\mathsf{KDer}(sk,\cdot)}(pk)$	$st \xleftarrow{\$} A_1^{\mathsf{KDer}(sk,\cdot)}(pk)$
$(m, t) \xleftarrow{\$} A_2(pk, st)$	$(m, t) \xleftarrow{\$} A_2(pk, st)$
$c \xleftarrow{\$} \mathsf{Enc}(pk, m)$	Let $f_1, \ldots, f_q$ be the queries made by A_1
$t' \xleftarrow{\$} A_3^{\mathcal{O}(sk,\cdot)}(pk, c, st)$	$t' \xleftarrow{\$} S^{\mathcal{O}'(sk,\cdot)}(pk, f_1(m), \ldots, f_q(m), st)$
If $t = t'$ return 1 else return 0	If $t = t'$ return 1 else return 0

where if mode = full then $\mathcal{O}(sk, \cdot) = \mathsf{KDer}(sk, \cdot)$ and for any f oracle $\mathcal{O}'(sk, f)$ returns $(sk_f, f(m))$ where $sk_f \xleftarrow{\$} \mathsf{KDer}(sk, f)$, and if mode = tna then $\mathcal{O}(sk, \cdot) = \mathcal{O}'(sk, \cdot) = \varepsilon$ (the empty oracle). We assume for simplicity that A_1's output (the state st) includes its oracle queries and the responses[6] and that $|m|$ in A_2's output depends only on k. Think of the string $t \in \{0,1\}^*$ in the output of A_2 as partial information on m. Note that in the above formalization of semantic security, even in the ideal experiment we run A_1 and A_2. A more standard formalization would have the simulator also run at these stages. However, we want to "bind" the simulator to making the same key derivation queries as the adversary.[7] Denote by $\Pr\left[\,\mathbf{Exp}^{\text{ss-real-mode}}_{\mathcal{FE},A}(k) = 1\,\right]$ the probability that the SS-REAL-MODE experiment outputs 1 and by $\Pr\left[\,\mathbf{Exp}^{\text{ss-ideal-mode}}_{\mathcal{FE},A,S}(k) = 1\,\right]$ the probability that the SS-IDEAL-MODE experiment outputs 1. Define

$$\mathbf{Adv}^{\text{ss-mode}}_{\mathcal{FE},A,S}(k) = \Pr\left[\,\mathbf{Exp}^{\text{ss-real-mode}}_{\mathcal{FE},A}(k) = 1\,\right] - \Pr\left[\,\mathbf{Exp}^{\text{ss-ideal-mode}}_{\mathcal{FE},A,S}(k) = 1\,\right] .$$

We say that $\mathcal{FE}$ is *SS-MODE secure* if for every PPT adversary A there exists a PPT simulator S such that $\mathbf{Adv}^{\text{ss-mode}}_{\mathcal{FE},A,S}(\cdot)$ is negligible.

3 Inequivalence of the Definitions in General

We investigate relations among the notions of security we introduced for FE. First, we note that when giving the adversary adaptive access to the token derivation oracle (i.e., what we call FULL security), one reason semantic security seems stronger than indistinguishability is that the simulator apparently needs to commit to a "dummy" ciphertext on which to run the adversary *before* knowing what values the challenge message should have when evaluated under the functions for which the adversary will later request tokens.

But we show that there is actually a more subtle reason for inequivalence of the definitions. In fact, we show that in general IND-FULL security does not even imply SS-*TNA* security. To show the separation we start with a IND-FULL secure functional encryption scheme for any class of functions $\mathcal{F}$ of a certain form. We then modify it to construct a new scheme that is still IND-FULL secure for $\mathcal{F}$ but *not* SS-TNA secure. We show the latter by presenting a concrete attack. We first describe a concept our counter-example scheme employs.

[6] We do not consider a notion of *function hiding* (cf. [SSW09]).

[7] On the other hand, in the case of FULL security this is not enforced by the definition. This was an oversight on our part that was not noticed until after seeing [BSW11], so we do not correct it here (all our results anyway concern the non-adaptive case). As in [BSW11] one could require that S and A_3 have the same query distribution. Other differences between the SS definition of [BSW11] and ours include: theirs considers only adaptive access to the key derivation oracle whereas ours distinguishes between adaptive and non-adaptive, and theirs considers the encryption of multiple messages whereas ours considers only a single message (in particular, the former is important for the proof of impossibility they give for meeting their notion without random oracles).

HIDDEN FUNCTIONS. Let $\mathcal{G} = \{g_k\}_{k\in\mathbb{N}}$ and $\mathcal{F} = \{f_k\}_{k\in\mathbb{N}}$ be families of functions on a common domain $D = D(k)$. We say $\mathcal{G}$ is *hidden* by $\mathcal{F}$ if any PPT adversary A on inputs $f_k, f_k(x)$ where $x \xleftarrow{\$} D$ outputs $g_k(x)$ with only negligible probability in k. Note that such functions can be constructed under standard assumptions; for example, let f_k be a one-way function applied to the first half of the bits of the input and let g_k just output these bits (that is, the first half of the bits of the input).[8] We say $\mathcal{F}$ and $\mathcal{G}$ are *isomorphic* if f_k and g_k are isomorphic for every k, meaning

$$f_k(d_1) = f_k(d_2) \Leftrightarrow g_k(d_1) = g_k(d_2) \ .$$

for all $d_1, d_2 \in D$. In other words, f_k and g_k have the same equality pattern across the domain. This is the case, for example, if f_k in the example above is an *injective* one-way function on the first half of the input bits. We supress dependence on k below for convenience, just talking of functions rather than function families.

THE COUNTER-EXAMPLE SCHEME. Let $\mathcal{AE}^* = (\mathsf{KDer}^*, \mathsf{Enc}^*, \mathsf{Dec}^*)$ be a (standard) public-key encryption scheme, and let $\mathcal{FE}' = (\mathsf{Setup}', \mathsf{KDer}', \mathsf{Enc}', \mathsf{Eval}')$ be a functional encryption scheme over message-space Σ for a class of functions $\mathcal{F} = \{f_1, \ldots, f_n\}$ satisfying the following: there is a function g on Σ such that the pointwise concatenated function[9] $f = f_1\| \ldots \|f_n$ is isomorphic to g and moreover g is hidden by f. (For simplicity, we assume here that n is polynomial in k. The counter-example can easily be extended to larger function sets by instead requiring the forgoing condition on some fixed *subset* of the f_i's.) Then we define a new functional encryption scheme $\mathcal{FE} = (\mathsf{Setup}, \mathsf{KDer}, \mathsf{Enc}, \mathsf{Eval})$ over Σ for $\mathcal{F}$ as follows.

- Setup on input 1^k first runs $(pk', sk') \xleftarrow{\$} \mathsf{Setup}'(1^k)$, and $(pk^*, sk^*) \xleftarrow{\$} \mathsf{KDer}^*(1^k)$. It then selects $w_1, \ldots, w_{n-1} \xleftarrow{\$} \{0,1\}^{|sk^*|}$ and computes $w_n \leftarrow sk^* \oplus w_1 \oplus \cdots \oplus w_{n-1}$. Finally, it returns master public key $pk = pk'\|pk^*$ and master secret key $sk = sk'\|w_1\| \ldots \|w_n$.
- KDer on input the master secret key $sk = sk'\|w_1\| \ldots \|w_k$ and a (description of a) function $f_i \in \mathcal{F}$ first runs $\mathsf{KDer}'_{sk'}(f_i)$ to obtain sk'_{f_i}. Then, it outputs $sk_{f_i} = sk'_{f_i}\|w_i$.
- Enc on input the master public key $pk = pk'\|pk^*$ and a message $m \in \Sigma$ first computes $c' \xleftarrow{\$} \mathsf{Enc}'(pk', m)$ and $c^* \xleftarrow{\$} \mathsf{Enc}^*(pk^*, g(m))$. It returns $c'\|c^*$.
- Eval on input a secret key $sk_{f_i} = sk'_{f_i}\|w_i$ and a ciphertext $c = c'\|c^*$ computes $d \leftarrow \mathsf{Eval}'(sk'_{f_i}, c')$, and outputs d.

Theorem 3.1 If $\mathcal{AE}^*$ is IND-CPA secure and $\mathcal{FE}'$ is IND-FULL secure for $\mathcal{F} = \{f_1, \ldots, f_n\}$ as above (i.e., where g is hidden by $f_1\| \ldots \|f_n$), then $\mathcal{FE}$ is also IND-FULL secure for $\mathcal{F}$. However, it is not SS-TNA secure.

Note that the assumptions of the theorem do not constitute any additional complexity assumptions beyond the (minimal) one of $\mathcal{FE}'$ being IND-FULL secure for $\mathcal{F}$, meaning based on the latter we can construct the other schemes and functions that are assumed.

We also remark that the separation also holds in the case of "selective-security," where the challenge messages are chosen up-front by the adversary, as considered in e.g. [BW07, KSW08]. It

[8] Indeed, a simpler example is to take f_k to be a one-way function and g_k to be the identity. However, in our counter-example this will prevent the adversary from even being able to *find* two messages that agree on f_k. We believe this points to a separate shortcoming of the IND definition.

[9] By pointwise concatenation $f\|g$ of functions f and g on a set D we mean that $f\|g(x) = f(x)\|g(x)$ for all $x \in D$.

also holds in the case of predicate encryption [BW07, KSW08], since we can take f_i for $1 \leq i \leq n$ to output the i-th bit of a function f such that g is hidden by f (i.e., a function can always be decomposed bit-wise into predicates).

Proof: (Sketch.) To see $\mathcal{FE}$ is IND-FULL secure for $\mathcal{F}$, first consider an adversary A that does not request tokens for all of $f_1, \ldots, f_n$. Then in addition to interacting with $\mathcal{FE}'$ in the IND-FULL experiment, the adversary is just given additional random strings when it requests tokens, which in particular are independent of b, so security of $\mathcal{FE}$ follows from that of $\mathcal{FE}'$. Now consider A that requests tokens for all of $f_1, \ldots, f_n$. In this case, in addition to interacting with $\mathcal{FE}'$ the adversary obtains $g(m_b)$ where m_b is the challenge message. But by the rules of the experiment we know that $f_1(m_0) \| \ldots \| f_n(m_0) = f_1(m_1) \| \ldots \| f_n(m_1)$ and thus by assumption $g(m_0) = g(m_1)$, meaning again this information is independent of b and so IND security of $\mathcal{FE}$ follows from that of $\mathcal{FE}'$.

To show that $\mathcal{FE}$ is not SS-TNA secure, we describe an SS-TNA adversary $B = (B_1, B_2, B_3)$ for which there is no simulator with comparable probability of guessing $t = t'$. Namely, B_1 requests evaluation tokens for all of $f_1, \ldots f_n$ and passes them along as the state, and B_2 chooses a random challenge message $m \in \Sigma$, sets $t \leftarrow g(m)$, and outputs (m, t). Then, by construction B_3 can always output $t = t'$ by decrypting the part of the challenge ciphertext formed by $\mathcal{AE}^*$ (note that B_3 makes no queries itself as required). However, a simulator who outputs $t = t'$ with non-negligible probability would contradict the fact that g is hidden by f since the simulator is not given any ciphertext but just the value of $f_1(m) \| \ldots \| f_n(m)$ (also pk and the evaluation tokens for $f_1, \ldots, f_n$, but a hidden function adversary can generate these itself). ∎

4 An Equivalence under Preimage Sampleability

We show that for token non-adaptive (TNA) security the counter-example in Section 3 is tight. Namely, we show an *equivalence* between indistinguishability and semantic-security under TNA security for what we call *preimage sampleable* (PS) schemes. Note that TNA security seems reasonable in practical applications where what tokens a party receives does not depend on the encrypted messages.

Preimage sampleability. Let $\mathcal{FE} = (\mathsf{Setup}, \mathsf{KDer}, \mathsf{Enc}, \mathsf{Eval})$ be a functional encryption scheme over message-space Σ for the class of functions $\mathcal{F}$. We call $\mathcal{FE}$ *preimage sampleable* (PS) if there is a PPT algorithm A such that, for every PPT algorithm B, the probability that the following experiment returns 0 is negligible in k:

Experiment $\mathbf{Exp}^{ps}_{\mathcal{FE},A,B}(k)$:
$(m, f_1, \ldots, f_\ell) \xleftarrow{\$} B(1^k)$
$m' \xleftarrow{\$} A(1^k, |m|, f_1(m), \ldots, f_\ell(m))$
If $|m| = |m'|$ and $f_i(m') = f_i(m)$ for all $1 \leq i \leq \ell$
  Then return 1
Else return 0

Above, we require that $m, m' \in \Sigma$ and $f_1, \ldots, f_\ell \in \mathcal{F}$.

We make a few remarks about our definition of preimage sampleability. First, we note that the inputs to A are always guaranteed to be consistent with *some* underlying m (and thus there is always at least one possible m' causing the PS experiment to return 1); on inputs that do

not satisfy this requirement we do not need the output of A to be defined. We also note that preimage sampleability as we have defined it is really a property of $\mathcal{F}$ and we sometimes refer to it as such. Finally, requiring that the inputs to A be generated by another PPT algorithm (rather than quantifying over all such inputs) is important to leave open the possibility of PS for some functionalities, such as 3-CNF formulae. (The latter point was brought to our attention by De Caro and Fiore [CF].)

In essence, we show that preimage sampleability provides a "test" of whether equivalence between the IND and SS definitions is maintained in the case of TNA security.

Theorem 4.1 Let $\mathcal{FE}$ be an PS functional encryption scheme. Then $\mathcal{FE}$ is SS-TNA secure if and only if it is IND-TNA secure.

Proof: (Sketch.) (SS-TNA $\Rightarrow$ IND-TNA) Suppose that $\mathcal{FE}$ is *not* IND-TNA secure, in particular let $A = (A_1, A_2)$ be a successful IND-TNA adversary against it. Consider SS-TNA adversary $B = (B_1, B_2, B_3)$ that works as follows. B_1 runs A_1 on pk (answering key-derivation queries using its own oracle) to receive (m_0, m_1). It then chooses $b \in \{0, 1\}$ at random and returns (m_b, b). B_3 runs A_2 on its input and outputs the result. Note that no SS-TNA simulator can output b with probability better than $1/2$ in the SS-TNA-IDEAL experiment because the simulator gets no information about b (since according to the rules of the IND-TNA experiment A may only makes token-derivation queries whose responses are independent of b, and the simulator makes no queries). So $\mathcal{FE}$ is not SS-TNA secure.

(IND-TNA $\Rightarrow$ SS-TNA) Now suppose $\mathcal{FE}$ *is* IND-TNA secure. Let $A = (A_1, A_2, A_3)$ be any SS-TNA adversary against $\mathcal{FE}$. We construct a simulator S with comparable success probability in the SS-IDEAL-TNA experiment to A in the SS-REAL-TNA experiment, which implies $\mathcal{FE}$ is SS-TNA secure. Simulator S works as follows: given queries $f_1, \ldots, f_q$ made by A_1 and their values $y_1, \ldots, y_q$ on the challenge message m, S will sample a "dummy" message $m' \in \Sigma$ such that $f_1(m') = y_1, \ldots, f_q(m') = y_q$ using the sampler A guaranteed by the definition of PS. (Here B in the definition of PS can be viewed as the entire experiment up to this point.) It runs A_3 on the encryption of m' and outputs the result. There are two cases:

- **Case 1:** $m = m'$ with overwhelming probability. Then A_3's success probability in the simulated environment remains negligibly different from the SS-TNA-REAL experiment.
- **Case 2:** $m \neq m'$ with non-negligible probability. Then if A_3's success probability also differs noticeably, we can construct a successful IND-TNA adversary $B = (B_1, B_2)$ against $\mathcal{FE}$, as follows. B_1 first runs A_1, A_2 on the appropriate inputs to receive (m, t). Let $f_1, \ldots, f_q$ be the queries made by A_1. Using the sampler guaranteed by the definition of PS, B_1 samples a message m' such that $f_1(m') = f_1(m), \ldots, f_q(m') = f_q(m)$, and returns (m, m', t) (i.e., m and m' are the challenge messages and t is the state). B_2 runs A_3 on pk, c from its input to recieve output t'; if $t' = t$ it returns 0, and otherwise 1. Note that for B to be successful it is important that $m \neq m'$, which holds with non-negligible probability in this case. This contradicts our initial assumption that $\mathcal{FE}$ is IND-TNA secure.

Thus in either case the success probability of S is close to that of A. ▌

It is interesting to note how the proof of the second implication accounts for the fact that IND-TNA may be "vacuously" satisfied when the adversary is not able to find two messages that agree on the

given functionality. Indeed, in this case, our simulator samples from the corresponding preimage set of size 1, and thus the simulation trivially works.

5 On Preimage Sampleability of Some Functionalities

We examine whether specific functionalities considered in the literature satisfy our PS condition. For inner-product predicates [KSW08, LOS+10, OT10], IBE [BF03], and PEKS [BCOP04, ABC+08], we show that the answer is "yes." On the other hand, for ABE [SW05, GPSW06, BSW07] it seems hard to show PS; we leave this as an open problem.

INNER-PRODUCTS. We first show that PS is satisfied by the important class of *inner-product predicates* realized in prior work [KSW08, LOS+10, OT10]. Hence, by Theorem 4.1, schemes in the literature for this functionality proven secure relative to the IND notion also meet SS, at least under non-adaptive access to the token-derivation oracle. Namely, consider the evaluation of inner products over $\mathbb{Z}_N$ for a composite N (of which it assumed hard to find a non-trivial factor; here N is generated by the PS experiment before being given to B, A). More formally, let $n \in \mathbb{N}$ be given and let N be such a composite. Let $\Sigma = \mathbb{Z}_N^n$ and define the associated class of *inner-product predicates* $\mathcal{P}_{iprod} = \{p_{\mathbf{x}} \mid \mathbf{x} \in \mathbb{Z}_N^n\}$ where $p_{\mathbf{x}}(\mathbf{y}) = 1$ if $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \cdot y_i = 0 \bmod N$, and 0 otherwise. (Note that in the terminology of [KSW08] we thus consider the "predicate-only" version of the scheme for simplicity.)

Proposition 5.1 The class $\mathcal{P}_{iprod}$ as defined above is PS if it is hard to find a non-trivial factor of N.

Proof: (Sketch.) We construct a PPT algorithm A that on input $(\mathbf{x}_1, y_1 = p_{\mathbf{x}_1}(\mathbf{m})), \ldots, (\mathbf{x}_r, y_r = p_{\mathbf{x}_r}(\mathbf{m}))$ for any polynomial $r = r(k)$ and any $p_{\mathbf{x}_1}, \ldots, p_{\mathbf{x}_r} \in \mathcal{P}_{iprod}$ and $\mathbf{m} \in \Sigma$, outputs a vector $\mathbf{m}'$ causing the PS experiment to return 1 with overwhelming probability. (Here we just refer to the probability over A's own coins.) Let I_b denote the set $\{i \in [r] \mid y_i = b\}$ for $b \in \{0, 1\}$, and let B be the $|I_1| \times n$ matrix where each row is a unique element of $\{\mathbf{x}_i \mid i \in I_1\}$.

Algorithm A works as follows. It first finds a basis $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_s\}$ for $\ker(B)$ in the space $\mathbb{Z}_N^n$. This is done by solving the homogeneous system of equations $B\mathbf{x} = \mathbf{0}$ using Gaussian elimination over $\mathbb{Z}_N$; while $\mathbb{Z}_N$ is not a field, if Gaussian elimination fails then A can find a non-trivial factor of N. It outputs a random $\mathbb{Z}_N$-combination of the vectors in W. That is, it outputs $\mathbf{m}' = r_1\mathbf{w}_1 + \ldots + r_s\mathbf{w}_s$ where each $r_i \in \mathbb{Z}_N$ for $1 \leq j \leq s$ is chosen independently at random.

For the analysis, we need to show that with overwhelming probability $\langle \mathbf{m}', \mathbf{x}_i \rangle = 0 \bmod N$ for all $i \in I_1$ and $\langle \mathbf{m}', \mathbf{x}_j \rangle \neq 0 \bmod N$ for all $j \in I_0$. The first part is clear by construction. For the second part, we first claim that for every *fixed* $j \in I_0$, the probablity over the choice of $r_1, \ldots, r_s \in \mathbb{Z}_N$ that $\langle \mathbf{m}', \mathbf{x}_j \rangle = 0 \bmod N$ is negligible. To see this, observe that there must be *some* (not necessarily unique) $\mathbf{w}_{i(j)} \in W$ such that $\langle \mathbf{x}_j, \mathbf{w}(j) \rangle \neq 0 \bmod N$, since otherwise there would be no $\mathbf{m}'$ causing the PS experiment to return 1. So, given any outcome of the r_i for $i \neq i(j)$ and assuming $\langle \mathbf{x}_j, \mathbf{w}(j) \rangle$ is not a zero-divisor (otherwise A can find a non-trivial factor of N), there is exactly *one* possible choice for $r_{i(j)}$ such that $\langle \mathbf{m}', \mathbf{x}_j \rangle = 0 \bmod N$. Now, by a union bound, the probability that $\langle \mathbf{m}', \mathbf{x}_j \rangle = 0 \bmod N$ for *any* $j \in I_0$ is negligible, which is what we needed to show. ▌

IBE AND PEKS. The functionalities for IBE [BF03] and PEKS [BCOP04, ABC+08] are also preimage sampleable. For example, in the case of IBE, given the functions and their values on the

underlying "message," there are two cases: if we know that $f_{ID}(ID'\|x) = x$ then there is only one possible preimage, namely $ID\|x$; otherwise, we can sample from the set of possible "messages" by choosing an identity other than those for which the adversary has requested secret keys and any payload (an analogous argument applies in the case of PEKS). We omit the formal statements. By Theorem 4.1, we conclude that such schemes in the literature proven secure under an IND notion also meet SS-TNA under this condition.

ATTRIBUTE-BASED ENCRYPTION. For the functionalities of ABE [SW05, GPSW06, BSW07], we do not know if PS holds. For example, consider the case of (anonymous) Fuzzy IBE [SW05, KSW08]. Namely, let U be a finite set and let Σ be the power-set of U, i.e., $\Sigma = \{S \mid S \subseteq U\}$. For $1 \leq d \leq |U|$ and $S, T \subseteq U$ define $p_{S,d}(T) = 1$ if $S \cap T \geq d$ and 0 otherwise. (As before, let us consider this "predicate-only" counterpart to Fuzzy IBE for simplicity.) Typically, one considers an FE scheme over Σ for the class $\mathcal{P}_d = \{p_{S,d} \mid S \subseteq U\}$ where d is fixed. To show PS, we would basically need to give an efficient algorithm that, given "good" sets $G_1, \ldots G_n \subseteq U$ and "bad sets" $B_1 \ldots, B_m \subseteq U$ for polynomials $n = n(k), m = m(k)$, as well as d such that $1 \leq d \leq |U|$, outputs a set $X \subseteq U$ such that $|X \cap G_i| \geq d$ for all $1 \leq i \leq n$ and $|X \cap B_j| < d$ for all $1 \leq j \leq m$. We are not sure if such an algorithm exists and leave this for future work.

Acknowledgements

We are very grateful to Alexandra Boldyreva for the initial conversations that led to this research, Mihir Bellare for discussions about the definitions and comments on the draft, and Nathan Chenette for discussions about preimage sampleability. We also thank Dario Fiore and Angelo De Caro for helpful feedback.

References

[ABC+08] Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, and Haixia Shi. Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions. *J. Cryptology*, 21(3):350–391, 2008.

[BCOP04] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public key encryption with keyword search. In *EUROCRYPT*, pages 506–522, 2004.

[BDPR98] Mihir Bellare, Anand Desai, David Pointcheval, and Phillip Rogaway. Relations among notions of security for public-key encryption schemes. In *CRYPTO*, pages 26–45, 1998.

[BF03] Dan Boneh and Matthew K. Franklin. Identity-based encryption from the weil pairing. *SIAM J. Comput.*, 32(3), 2003.

[BSW07] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In *IEEE Symposium on Security and Privacy*, pages 321–334, 2007.

[BSW11] Dan Boneh, Amit Sahai, and Brent Waters. Functional encryption: Definitions and challenges. In *TCC*, 2011.

[BW07] Dan Boneh and Brent Waters. Conjunctive, subset, and range queries on encrypted data. In *TCC*, pages 535–554, 2007.

[CF] Angelo De Caro and Dario Fiore. Personal correspondence, 2010.

[CK10] Melissa Chase and Seny Kamara. Structured encryption and controlled disclosure. In *ASIACRYPT*, pages 577–594, 2010.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[GPSW06] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In *ACM Conference on Computer and Communications Security*, pages 89–98, 2006.

[KSW08] Jonathan Katz, Amit Sahai, and Brent Waters. Predicate encryption supporting disjunctions, polynomial equations, and inner products. In *EUROCRYPT*, pages 146–162, 2008.

[LOS+10] Allison B. Lewko, Tatsuaki Okamoto, Amit Sahai, Katsuyuki Takashima, and Brent Waters. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. In *EUROCRYPT*, pages 62–91, 2010.

[MRS88] Silvio Micali, Charles Rackoff, and Bob Sloan. The notion of security for probabilistic cryptosystems. *SIAM J. Comput.*, 17(2), 1988.

[Nie02] Jesper Buus Nielsen. Separating random oracle proofs from complexity theoretic proofs: The non-committing encryption case. In *CRYPTO*, pages 111–126, 2002.

[OT10] Tatsuaki Okamoto and Katsuyuki Takashima. Fully secure functional encryption with general relations from the decisional linear assumption. In *CRYPTO*, pages 191–208, 2010.

[SSW09] Emily Shen, Elaine Shi, and Brent Waters. Predicate privacy in encryption systems. In *TCC*, pages 457–473, 2009.

[SW05] Amit Sahai and Brent Waters. Fuzzy identity-based encryption. In *EUROCRYPT*, pages 457–473, 2005.

[Wat] Brent Waters. Functional encryption: Beyond public-key cryptography. Presentation available from `http://userweb.cs.utexas.edu/bwaters`.