Two Faces of Blindness[⋆]

Jaap-Henk Hoepman[1,2]

[1] Radboud University Nijmegen, Email: jhh@cs.ru.nl
[2] Karlstad University
[3] University of Groningen

Abstract. Blind signatures are a decades-old privacy enhancing technology. It is not always clearly understood that blind signatures actually possess two separate properties: the intuitive understanding that the message to be signed is hidden from the signer, and the fact that the resulting signature is unlinkable (meaning that the signer cannot later tell in which session it created a particular signature). The question is: how exactly should these properties be defined, and can they be defined in a natural way such that they are mutually independent yet together imply blindness?

In this paper we study this question, present formal definitions for *message indistinguishability* and *signature unlinkability* (and a few more related ones), and study their relationships. We show that these two properties are indeed mutually independent. Unfortunately their union is not equivalent to blindness in what appear to be only pathological cases.

1 Introduction

David Chaum introduced blind signatures almost four decades ago [6], as the fundamental building block to implement a form of untraceable digital cash. His proposal was to represent each digital coin as a unique serial number blindly signed by the issuing bank. The unique serial number embedded in the coin would prevent double spending, while the blind signature over the coin would guarantee both *untraceability* (by not knowing which coin was signed) and *unforgeability* (by signing the coins in the first place).

Chaum explained blind signatures intuitively by showing how a blind signature could be implemented in a traditional, non digital, setting using carbon paper inside paper envelopes. To obtain a blind signature on a secret message, a user could send the message inside a sealed envelope to the

[⋆] This paper significantly extends ideas that originally were mentioned in [7]. Version: Tue Oct 18 19:40:10 2022 +0200 / DESI-final / two-types-of-blind-sigs.tex signer, with the inside of the envelope covered with carbon paper. The carbon paper ensures that if the signer signs the envelope from the outside, the carbon paper transfers this signature to the secret message inside the envelope. When the signer returns the still sealed envelope (proving it didn't see the message) all the user needs to do is to open the envelope to obtain the blindly signed message.

This intuitive explanation clearly shows that the message stays hidden from the signer. But this by itself is not enough to prevent a bank from tracing a digital coin signed this way, even if it prevents the bank from learning its serial number. In fact, if the bank signs each envelope in a slightly different way, and remembers which way of signing it used to sign each envelope, it can link actual signatures on messages to the particular envelope on which it put the exact same signature. In other words, in order to guarantee all the desired security and privacy properties, blind signatures need to guarantee the following two separate properties.

"Hiding the message"
The message to be signed is hidden from the signer.

Signature unlinkability
Given a final blind signature on a message, the signer cannot determine when it generated that particular signature.

Perhaps due to Chaum's metaphor, blind signatures have always informally been explained as signatures where the message to be signed is hidden from the signer. But as the above example shows, blind signatures need to guarantee two separate 'faces of blindness'. The question is: how exactly should these properties be defined, and can they be defined in a natural way such that they are mutually independent yet together imply blindness?

Although in the particular case of signing digital coins signing messages without knowing their contents is a desirable feature, in general this is irresponsible: who would sign a contract without knowing its terms? Therefore, in many applications *partially* blind signatures, where the signer may need to know (at least part of) the message before signing it, do serve an important purpose. Such partially blind signatures have been introduced by Abe and Fujisaki [1], and have applications in scenarios where a user wants to prove that a certain condition has been met, without revealing when or where that condition was met. Blind signatures can for example be used to issue a unique and unforgeable token or receipt whenever a user has performed a certain action (like paying a bill, visiting a checkpoint, entering or leaving a certain location, completing some task, or satisfying any other predetermined requirement). This token can then later be used to prove that this particular action was performed or requirement was satisfied. This approach has been used, for example, to construct a privacy friendly form of ticketing for public transport [7]. Blind signatures have also been used to implement attribute based credentials [4, 5, 8].

In this paper we explore the different faces of blindness in depth, in the more general setting of partially blind signatures. We note that our results also apply to normal blind signatures as such signatures are equivalent to partially blind signatures where the public message equals the empty string. In a way this paper is a dual to the paper of Schröder and Unruh [15] that reexamines the definition of security of blind signature schemes, discovering that the messages and their resulting signatures have some independent influence on the overall security of the scheme.

We first define partially blind signatures and their completeness and unforgeability properties in section 2. We then study the two faces of blindness (*message indistinguishability* and *signature unlinkability*) and their relationships in section 3. This section also discusses message hiding, and why message indistinguishability is the more appropriate notion to study in this context. We show that message indistinguishability and signature unlinkability are both implied by a partially blind signature scheme, but that they are indeed two separate notions (in the sense that there are signature schemes that satisfy one of the two requirements, but bot both). Unfortunately, in pathological cases the union of these two properties does not imply blindness. We summarise and discuss our results in figure 1 and section 4. Figure 1 is also useful as a 'cheat-sheet' to keep track of the different properties defined throughout the paper.

2 Completeness and unforgeability

We start with the basics: the definition of completeness and unforgeability of (partially blind) signatures. We follow the framework for defining blind signatures provided by Juels *et al.* [10] and generalised and refined for partially blind signatures by Abe and Okamoto [2, 12, 13]). In this setting a (partially blind) signature scheme is defined as follows (where λ is the security parameter of the scheme).[4]

Definition 2.1 (signature scheme). *A* signature scheme Σ *consists of four probabilistic polynomial-time algorithms* $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$.

[4] The following definitions actually apply to arbitrary signature schemes, except that the artificial distinction between public and private messages is only relevant for partially blind signatures.

- *$\mathscr{G}$ takes security parameter 1^λ as input, and returns a secret key sk (to be given to the signer only) and a corresponding public key PK (known to all parties in the system).*
- *$\mathscr{S}$ and $\mathscr{U}$ are in fact interactive algorithms where signer $\mathscr{S}$ has private input sk and public input the public message $\overline{m}$ (with length polynomial in the security parameter λ), while user $\mathscr{U}$ has private input message m (also with length polynomial in the security parameter λ) and public input PK and $\overline{m}$. $\mathscr{S}$ and $\mathscr{U}$ interact with each other over a public communication channel. After the interaction, $\mathscr{S}$ outputs either* **success** *or* **fail**, *and $\mathscr{U}$ outputs either a signature σ or $\perp$. $\mathscr{U}$'s output is private. $\mathscr{S}$'s output is public.*
- *$\mathscr{V}$ takes as input a public key PK, public message $\overline{m}$, a message m and a signature σ, and outputs either* **accept** *or* **reject**. *This verification can be performed by any party.*

We write $out_{\mathscr{S}} \leftarrow \mathscr{S}(\mathrm{sk}, \overline{m}) \Leftrightarrow \mathscr{U}(\mathrm{PK}, \overline{m}, m) \rightarrow out_{\mathscr{U}}$ for an interaction between a signer and a user with the specified inputs, with $out_{\mathscr{S}}$ as the output of the signer and $out_{\mathscr{U}}$ as the output of the user.

Definition 2.2 (Completeness). *A signature scheme $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ is* complete *when for every interaction*

$$\mathbf{success} \leftarrow \mathscr{S}(sk, \overline{m}) \Leftrightarrow \mathscr{U}(PK, \overline{m}, m) \rightarrow \sigma$$
$$\textit{such that } \mathscr{V}(PK, \overline{m}, m, \sigma) = \mathbf{accept}$$

holds with overwhelming probability (i.e., with probability $1 - 2^{-\lambda}$), where this probability is computed over the private coin-flips of $\mathscr{G}, \mathscr{S}, \mathscr{U}$ and $\mathscr{V}$.

We return to this somewhat peculiar definition of completeness (that subsumes correctness) in the next section.

We now define the unforgeability property.

Definition 2.3 (Unforgeability). *Let $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ be a signature scheme and consider the following game between an adversarial user $\mathscr{U}^*$ and a honest signer $\mathscr{S}$ and honest verifier $\mathscr{V}$.*

1. *Run $\mathscr{G}(1^\lambda)$ to generate sk and PK. Give sk,PK to $\mathscr{S}$ and PK to $\mathscr{U}^*$.*
2. *Let $\mathscr{U}^*$ engage in polynomially (in λ) many adaptive, parallel and arbitrarily interleaved interactions with polynomially many copies of the signer $\mathscr{S}$ (knowing sk). Let j be the number of such interactions that return* **success** *for the signer.*
3. *Let $\mathscr{U}^*$ return a list of k signatures $\sigma_1, \ldots, \sigma_k$ for k distinct combinations of public messages and private messages $(\overline{m}_i, m_1), \ldots, (\overline{m}_k, m_k)$ such that $\mathscr{V}(PK, \overline{m}_i, m_i, \sigma_i) =$* **accept** *for all $i \in \{1, \ldots, k\}$.*

Adversary $\mathcal{U}^*$ wins this game whenever $k > j$.

The signature scheme is unforgeable *when every possible adversary $\mathcal{U}^*$ wins this game with at most negligible probability (i.e., probability $2^{-\lambda}$), where this probability is computed over the private coin-flips of $\mathcal{G}$, $\mathcal{U}^*$, $\mathcal{V}$ and all signers $\mathcal{S}$.*

3 The two faces of blindness

With the above definitions for a correct and unforgeable signature scheme given we are now ready to study the two different faces of blindness of such signature schemes.

We start with the definition of blindness itself. After that we study message indistinguishability in section 3.2. This notion is somewhat stronger than message hiding (discussed in section 3.3). We finish with the definition of signature unlinkability in section 3.4. It turns out that it is more appropriate to focus on message indistinguishability rather than message hiding, because the latter notion is actually implied by signature unlinkability. Throughout this section we establish relationships between the different notions we define.

3.1 Blindness

The following definition of partial blindness is due to Abe and Okamoto [2, 13] that extends the original defintion of blind signatures from Juels *et al.* [10] by allowing part of the message to be signed to be public.

Definition 3.1 (Blindness). *Consider a signature scheme $\langle \mathcal{G}, \mathcal{S}, \mathcal{U}, \mathcal{V} \rangle$ and the following game between an adversarial signer $\mathcal{S}^*$ and two honest users $\mathcal{U}_0$ and $\mathcal{U}_1$, mediated by a challenger.*

1. *Run $\mathcal{G}(1^\lambda)$ to generate sk and PK. Give sk, PK to $\mathcal{S}^*$.*
2. *Adversary $\mathcal{S}^*$ outputs PK, two private messages[5] m_0, m_1, and public message $\overline{m}$, and gives them to the challenger.*
3. *The challenger randomly selects $b \in \{0, 1\}$ and sets $\bar{b} = 1 - b$. It sets up user $\mathcal{U}_0$ with input $(PK, \overline{m}, m_b)$ and user $\mathcal{U}_1$ with input $(PK, \overline{m}, m_{\bar{b}})$.*
4. *$\mathcal{S}^*$ is given oracle access to each of these users to engage in the blind signature protocol with each of them, mediated by the challenger.[6]*

[5] Observe that these two messages are not required to be distinct.

[6] Observe that in this game (and the ones that follow) we do not need to allow the adversary to engage in polynomial many runs, for the simple reason that the adversary is now the signer, who given the private key can simulate all runs for himself.

5. *Let σ_b be the result returned by $\mathscr{U}_0$ and $\sigma_{\bar{b}}$ be the result returned by $\mathscr{U}_1$. If both signatures are valid, the challenger gives (σ_0, σ_1) to $\mathscr{S}^*$, in that fixed order. Give $\perp$ to $\mathscr{S}^*$ otherwise.*
6. *$\mathscr{S}^*$ outputs $b' \in \{0,1\}$.*

Adversary $\mathscr{S}^*$ wins this game whenever $b' = b$. The signature scheme is blind *when every possible adversary $\mathscr{S}^*$ wins this game with at most negligible advantage (i.e., probability $1/2 \pm 2^{-\lambda}$), where the probability is computed over the coin-flips of $\mathscr{S}^*$ and the private coin-flips of $\mathscr{U}_0$ and $\mathscr{U}_1$.*

Note that in this definition, as well as the ones that follow, we assume that the adversarial signer knows which of the users ($\mathscr{U}_0$ or $\mathscr{U}_1$) it is interacting with during the protocol.

The above definition is taken from [13], which differs in one significant aspect from [12] (the published conference version that precedes the full paper [13]) as follows. Step 5 in the game above originally read:

5' Let σ_b be the result returned by $\mathscr{U}_0$ and $\sigma_{\bar{b}}$ be the result returned by $\mathscr{U}_1$. If both signatures are valid, the challenger gives $(\overline{m}, m_b, \sigma_b)$ and $(\overline{m}, m_{\bar{b}}, \sigma_{\bar{b}})$ to $\mathscr{S}^*$ in arbitrary order. If only one of the signatures is valid, the challenger gives that signature and the corresponding message to $\mathscr{S}^*$. Give $\perp$ to $\mathscr{S}^*$ otherwise.

In other words: the original game allows that even if only one of the signatures is valid, the challenger gives that signature and the corresponding message to $\mathscr{S}^*$. This leaves a blind signature scheme open to the following generic attack.

1. Adversary $\mathscr{S}^*$ outputs PK and two private messages m_0, m_1, and public message $\overline{m}$, and gives them to the challenger.
2. The challenger randomly selects $b \in \{0,1\}$ and sets $\bar{b} = 1 - b$. It sets up user $\mathscr{U}_0$ with input (PK, $\overline{m}, m_b$) and user $\mathscr{U}_1$ with input (PK, $\overline{m}, m_{\bar{b}}$).
3. $\mathscr{S}^*$ engages in the blind signature protocol, *but only with $\mathscr{U}_0$*. It aborts its interaction with $\mathscr{U}_1$ which therefore returns $\perp$. (Note: $\mathscr{U}_1$ can also return a random value, but definitely not a valid signature as this requires the cooperation of $\mathscr{S}^*$, so this is easily detected in the next step.)
4. Let σ_b be the result returned by $\mathscr{U}_0$. As the other signature equals $\perp$ the challenger therefore gives (m_b, σ_b) to $\mathscr{S}^*$ as its challenge.
5. This is no game for $\mathscr{S}^*$: using its knowledge of m_0 and m_1 it quickly sees which of the two was given to $\mathscr{U}_0$ to sign. $\mathscr{S}^*$ outputs $b \in \{0,1\}$ and wins.

Clearly this is not desirable, which probably explains why the definition is amended in the full paper.

3.2 Message indistinguishability

We now turn our attention to the message indistinguishability property, stating that the adversary cannot distinguish which of two known messages it is actually asked to sign by a user.

Definition 3.2 (Message indistinguishability). *Let $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ be a signature scheme and consider the following game between an adversarial signer $\mathscr{S}^*$ and a honest user $\mathscr{U}$, mediated by a challenger.*

1. *Run $\mathscr{G}(1^\lambda)$ to generate sk and PK. Give sk,PK to $\mathscr{S}^*$.*
2. *Adversary $\mathscr{S}^*$ outputs PK and two private messages m_0, m_1, and public message $\overline{m}$, and gives them to the challenger.*
3. *The challenger randomly selects $b \in \{0, 1\}$. It sets up user $\mathscr{U}$ with input PK, $\overline{m}, m_b$.*
4. *$\mathscr{S}^*$ is given oracle access to the user to engage in the blind signature protocol with it, mediated by the challenger.*
5. *Let σ be the result returned by $\mathscr{U}$. This is hidden from $\mathscr{S}^*$.*[7]
6. *$\mathscr{S}^*$ outputs $b' \in \{0, 1\}$.*

Adversary $\mathscr{S}^*$ wins this game whenever $b' = b$.

The signature scheme is message indistinguishable *when every possible adversary $\mathscr{S}^*$ wins this game with at most negligible advantage (i.e., probability $1/2 \pm 2^{-\lambda}$), where the probability is computed over the coin-flips of $\mathscr{S}^*$ and the private coin-flips of $\mathscr{U}$.*

We first offer an example of a signature scheme that is message indistinguishable, as this is useful in the proofs that follow. This signature scheme requires a semantically secure encryption scheme $\{\}_k$ that satisfies the following property.

Property 3.1. Given c, m and k such that $c = \{m\}_k$, the probability to find $m' \neq m$ and a potentially different key k_x such that $c = \{m'\}_{k_x}$ is negligible.

One might think that an authenticated encryption scheme perhaps fits the bill [3]. Unfortunately this is in general not the case.[8] Luckily, a special

[7] Note that we cannot give σ to $\mathscr{S}^*$, as this would allow $\mathscr{S}^*$ to easily test which message was signed using the public verification function. The adversarial signer *must* derive information about the message signed only from the transcript of the protocol run.

[8] In fact, the general Encrypt-then-MAC approach using encryption key k_E and tagging key k_T (that returns the pair (c, t) as ciphertext where $c = E_{k_E}(m)$ and $t = S_{k_T}(c)$) does not work because k_E and k_T are unrelated. Given (c, t) as an encryption of m against key k_E, k_T, we can pick an arbitrary key $k_{E'}$, use it to decrypt c to obtain m' such that $c = E_{k_{E'}}(m')$ and leave the tag alone. Then (c, t) is a valid encryption for m' as well (based on keys $k_{E'}$ and k_T).

mode of authenticated encryption called CCM (that combines CTR encryption with a CBC-MAC *using the same key* k) satisfies this property. CCM is a stream cipher that roughly works as follows (see [9] for details).

- Let $E_k()$ be a pseudo-random function (it could be a block cipher or a hash function keyed by k).
- Let m be a message whose length is a multiple of the block length of this underlying block cipher, and write $m = m_1 \| \dots \| m_z$.
- Compute the tag t for message m by using $E_k()$ in CBC mode: define $t_1 = E_k(m_1)$, let $t_{i+1} = E_k(m_{i+1} \oplus t_i)$ and let $t = t_z$. We write $t = T_k(m)$ Again (for simplicity) tags are assumed to be exactly as long as a single block.
- Compute the key stream blocks A_i by encrypting a counter with k, i.e., $A_i = E_k(i)$.
- The full CCM ciphertext is obtained by XOR-ing $m \| t$ with $A_0 \| \dots \| A_z$.

CCM is known to be semantically secure [9]. We show it also satisfies property 3.1.

Lemma 3.1. *Let $\{m\}_k$ be the CCM authenticated encryption scheme described above. Such a scheme satisfies property 3.1.*

Proof. Suppose we have $c = (m \| t) \oplus (A_0 \| \dots \| A_z)$, where $A_i = E_k(i)$ and $t = T_k(m)$. Let $c = c_0 \| \dots \| c_z$. If we focus on the tag part, then to break the property we need to find m' and k_x such that $c = (m' \| t') \oplus (A'_0 \| \dots \| A'_z)$, where $A'_i = E_{k_x}(i)$ and $t' = T_{k_x}(m')$. This entails finding m' and k_x such that $c_z = t' \oplus A'_z = T_{k_x}(m') \oplus E_{k_x}(z)$. In this equation c_z and z are fixed. The adversary is free to choose k_x but this fixes m' as well as it needs to match c when xor-ed with $(A'_0 \| \dots \| A'_{z-1})$. If we model the pseudo-random function $E_k()$ as a random oracle [11], it is extremely unlikely that it is possible to meet these constraints: for every possible choice of k there is exactly one possible mapping of the random oracle for $E_{k_x}(z)$ that satisfies the equation, which only happens with negligible probability.

Construction 3.1 (Message indistinguishable signature scheme) *Let $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ be any ordinary unforgeable and complete signature scheme (where $\mathscr{U}$ submits the message m to be signed in plaintext to $\mathscr{S}$; we are abusing notation somewhat). Let $\{m\}_{k_U}$ be the CCM authenticated encryption scheme discussed above.*

Define the message indistinguishable signature scheme $\Sigma' = \langle \mathscr{G}', \mathscr{S}', \mathscr{U}', \mathscr{V}' \rangle$ as follows. $\mathscr{G}'$ equals $\mathscr{G}$ creating signing key $k_{\mathscr{S}}$ and verification key $K_{\mathscr{S}}$.

User $\mathscr{U}'$, before submitting a message m to be signed, generates a key $k_{\mathscr{U}'}$. It encrypts the message m as $c = \{m\}_{k_{\mathscr{U}'}}$ using the CCM encryption scheme and *sends this to the signer who creates the intermediate signature $\sigma' = [c \,\|\, \overline{m}]_{k_{\mathscr{S}}}$ (using its knowledge of $k_{\mathscr{S}}$ and public parameter $\overline{m}$). It returns this to $\mathscr{U}$ who adds $k_{\mathscr{U}'}$ to create the final signature $\sigma = (\sigma', k_{\mathscr{U}'})$. $\mathscr{U}'$ outputs σ and $\mathscr{S}'$ outputs* **success**. *This describes $\mathscr{S}'$ and $\mathscr{U}'$.*

Signature verification $\mathscr{V}'$ then runs as follows. Given $K_{\mathscr{S}}$, σ, $\overline{m}$, and m, the verifier first uses $k_{\mathscr{U}}$ embedded in σ to reconstruct $c = \{m\}_{k_{\mathscr{U}}}$. It then verifies that indeed $\sigma = [c \,\|\, \overline{m}]_{k_{\mathscr{S}}}$ using the public key $K_{\mathscr{S}}$ and the original signature verification function $\mathscr{V}$.

Lemma 3.2. *The signature scheme in construction 3.1 is message indistinguishable according to definition 3.2.*

Proof. The construction matches the (syntactic) constraints of definition 2.1, and it is easily seen to be complete as defined in 2.2.

We rely on property 3.1 to prove unforgeability (definition 2.3). If the blind signature scheme would be forgeable, a user $\mathscr{U}^*$ would be able to return k signatures $\sigma_1, \ldots, \sigma_k$ for k distinct messages $(\overline{m}_1, m_1), \ldots, (\overline{m}_k, m_k)$ such that $\mathscr{V}(\mathrm{PK}, \overline{m}_i, m_i, \sigma_i) =$ **accept** for all $i \in \{1, \ldots, k\}$, when given only $j < k$ such message/signature pairs. By definition, the underlying standard signature scheme is not forgeable. By the pigeonhole principle then there should be two signatures $\sigma_i = (\sigma'_i, k_i)$ and $\sigma_j = (\sigma'_j, k_j)$ such that σ'_i and σ'_j are signatures over the equal strings $c_i \,\|\, \overline{m}_i$ and $c_j \,\|\, \overline{m}_j$. Then $\overline{m}_i = \overline{m}_j$ and $c_i = \{m_i\}_{k_i} = \{m_j\}_{k_j} = c_j$ while $(\overline{m}_i, m_i) \neq (\overline{m}_j, m_j)$ by assumption. This contradicts property 3.1.

Because the encryption scheme is semantically secure, this signature scheme is message indistinguishable according to definition 3.2.

We first show that blindness implies message indistinguishability.

Theorem 3.2. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is blind according to definition 3.1. Then Σ is message indistinguishable according to definition 3.2.*

Proof. Intuitively the argument runs as follows. Because the signer knows that b selects which message user $\mathscr{U}_0$ will offer for signing, if the signature scheme were not message indistinguishable, the signer could trivially guess b correctly (even when not given m_b). The formal proof requires a bit more work.

Suppose not. So there is an adversarial signer $\mathscr{S}^*$ for the game defined in definition 3.2. We turn it into an adversarial signer $\mathscr{S}^{**}$ for the game defined in definition 3.1 as follows.

1. $\mathscr{S}^{**}$ starts $\mathscr{S}^*$, which returns PK and two private messages m_0, m_1, and public message $\overline{m}$.

2. $\mathscr{S}^{**}$ forwards these to the challenger from definition 3.1.
3. Let this challenger randomly select $b \in \{0,1\}$, set $\bar{b} = 1 - b$, giving user $\mathscr{U}_0$ the input $(\mathrm{PK}, \overline{m}, m_b)$ and user $\mathscr{U}_1$ the input $(\mathrm{PK}, \overline{m}, m_{\bar{b}})$.
4. Set up both users to be ready to engage with $\mathscr{S}^{**}$ in the blind signature protocol (according to the game defined in 3.1).
5. $\mathscr{S}^{**}$ is merely a mediator now, relaying messages between the users and $\mathscr{S}^{*}$. It actually runs the interactive blind signing protocol only between user $\mathscr{U}_0$ and $\mathscr{S}^{*}$. (It aborts the other instance.) Observe how this corresponds to the challenge that $\mathscr{S}^{*}$ is supposed to get according to definition 3.2.
6. Let σ_b be the result returned by $\mathscr{U}_0$. (The other user returns $\perp$.)
7. Because one of the signatures fails to be created, according to the blindness game defined for definition 3.1, the challenger gives $\perp$ to $\mathscr{S}^{**}$, who simply discards it.
8. $\mathscr{S}^{*}$ outputs $b' \in \{0,1\}$, which $\mathscr{S}^{**}$ forwards as its own output for this challenge.

The output b' of $\mathscr{S}^{*}$ corresponds to the challenge $\mathscr{U}_0, \mathrm{PK}, \overline{m}, m_b$. If $b = b'$, then by construction b' is also the correct response to the challenge given to $\mathscr{S}^{**}$. This shows that advantage of $\mathscr{S}^{**}$ the same of that of $\mathscr{S}^{*}$, i.e., non-negligible, contradicting the premise of the theorem.

The converse does not hold however: there are message indistinguishable signature schemes that are not blind as the following theorem demonstrates. This shows that message indistinguishability is a strictly weaker notion.

Theorem 3.3. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is message indistinguishable according to definition 3.2. This does* not *imply that Σ is blind according to definition 3.1.*

Proof. Let Σ be the signature scheme from construction 3.1. This is message indistinguishable according to lemma 3.2.

Clearly this signature scheme is not really blind: a malicious signer can record for each run the signature σ' it generated. It can then always win the game in definition 3.1: it now knows the σ'_b it created while interacting with $\mathscr{U}_0$, which it can match to $(\sigma_0, \sigma_1) = ((\sigma'_0, k_{\mathscr{U}_b}), (\sigma'_1, k_{\mathscr{U}_{\bar{b}}}))$ (where σ'_b is the signature over m_b). This reveals b.

We conclude that message indistinguishability does not imply blindness, and thus the theorem follows.

3.3 Message hiding

Message indistinguishability is a very strong property (it is in fact very similar to semantic security definitions for encryption schemes [11]), but perhaps this property is somewhat counter intuitive and perhaps even stronger than needed for the typical scenario where blind signatures are used: there we typically want to prevent the signer from *learning* a *random* message (think a random sequence number) *someone else* submits for signing. This notion is captured in the following definition of message hiding.

Definition 3.3 (Message hiding, strong version). *Let $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ be a signature scheme and consider the following game between an adversarial signer $\mathscr{S}^*$ and an honest user $\mathscr{U}$, mediated by a challenger.*

1. *Run $\mathscr{G}(1^\lambda)$ to generate sk and PK. Give sk, PK to $\mathscr{S}^*$.*
2. *Adversary $\mathscr{S}^*$ outputs PK and public message $\overline{m}$, and gives them to the challenger.*
3. *The challenger randomly selects a private message $m \in \{0,1\}^\lambda$, and sets up an instance of a user $\mathscr{U}$ with input PK, $\overline{m}, m$.*
4. *$\mathscr{S}^*$ is given oracle access to user $\mathscr{U}$ to engage in the blind signature protocol with it, mediated by the challenger.*
5. *Let σ be the signature returned by $\mathscr{U}$. The challenger gives σ to $\mathscr{S}^*$.*
6. *$\mathscr{S}^*$ outputs $m' \in \{0,1\}^\lambda$.*

Adversary $\mathscr{S}^*$ wins this game whenever $m' = m$.

The signature scheme is message hiding *when every possible adversary $\mathscr{S}^*$ wins this game with at most negligible probability (i.e., probability at most $2^{-\lambda}$), where the probability is computed over the coin-flips of $\mathscr{S}^*$ and the private coin-flips of $\mathscr{U}$.*

Blind signature schemes that only offer message hiding are for instance used in the Idemix attribute based credential system to hide the master secret m_1 from the credential issuer [8]. A trivial implementation of such a blind signature scheme in the random oracle model would be one where the message m to be signed is first hashed using a cryptographic hash function h and subsequently sending the resulting hash $h(m)$ to the signer to be signed with an arbitrary traditional (non-blind) signature scheme.[9]

This shows that message hiding is a strictly weaker notion than (general) blindness. But does message indistinguishability imply message hiding, or the other way around? In fact not when we define message hiding as above.

Theorem 3.4. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is message hiding according to definition 3.3. This does* not *imply that Σ is message indistinguishable according to definition 3.2.*

[9] Idemix uses a Pedersen style commitment [14] instead of a hash function to blind the secret master key. This makes it possible to prove knowledge of this secret in a zero knowledge proof later.

Proof. Consider the basic message hiding signature scheme above. Let h be a hash function modelled as a random oracle. This guarantees that no adversary is able to recover m given $h(m)$.

Let the signer use an ordinary signature scheme with signing key k_S and verification key K_S to compute the signature σ on a string s as $[s]_{k_S}$. A message hiding signature scheme is one where the user, wishing to compute a signature on a public message $\overline{m}$ and a private message m computes $\overline{m} \,||\, h(m)$ and sends this to the signer to sign. The signature then equals $[\overline{m} \,||\, h(m)]_{k_S}$. To verify such a signature, the verifier is given $\overline{m}$ and m, computes $\overline{m} \,||\, h(m)$ and uses checks the signature σ using the underlying traditional signature verification function.

The construction matches the (syntactic) constraints of definition 2.1, and it is easily seen to be complete as defined in 2.2.

The construction is also (strongly) message hiding according to definition 3.3. Suppose the challenger returns a signature σ after the query phase. If the adversary is able to successfully guess m' such that $\sigma = [\overline{m} \,||\, h(m')]_{k_S}$ then this essentially means the adversary was able to compute $m' = m$ while observing the hashes $h(m)$ sent during the signing process. This is contrary to the assumption on h.

The thus constructed signature scheme is clearly not message indistinguishable according to definition 3.2. If the adversary selects $\overline{m}, m_0, m_1$ and receives $\overline{m} \,||\, h(m_b)$ for signing, it easily checks which of the two m_0 and m_1 matches $h(m_b)$ to correctly guess b.

Theorem 3.5. *Consider a signature scheme $\Sigma = \langle \mathcal{G}, \mathcal{S}, \mathcal{U}, \mathcal{V} \rangle$ that is message indistinguishable according to definition 3.2. This does* not *imply that Σ is message hiding according to definition 3.3.*

Proof. Let Σ be the message indistinguishable signature scheme from construction 3.1. Suppose we tweak it a bit such that the signature returned by the user equals $\sigma = (\sigma', k_{\mathcal{U}}, \overline{m}, m)$. This tweak does not affect message indistinguishability, for in that game σ is not given to the adversary as part of the challenge. However, in the message hiding game as defined in definition 3.3, the adversary *does* get σ and thus trivially wins that game. The result follows.

So message indistinguishability and strong message hiding are incomparable notions. However, a weaker notion of message hiding (that does not give the adversary access to the generated signatures) does follow from message indistinguishability. For that we have to weaken the definition a bit by *not* giving the adversarial signer the set of final signatures obtained by the user(s). The formal definition is as follows.

Definition 3.4 (Message hiding). *Let $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ be a signature scheme and consider the following game between an adversarial signer $\mathscr{S}^*$ and a honest user $\mathscr{U}$, mediated by a challenger.*

1. *Run $\mathscr{G}(1^\lambda)$ to generate sk and PK. Give sk,PK to $\mathscr{S}^*$.*
2. *Adversary $\mathscr{S}^*$ outputs PK and public message $\overline{m}$, and gives them to the challenger.*
3. *The challenger randomly selects a private message $m \in \{0,1\}^\lambda$, and sets up an instance of a user $\mathscr{U}$ with input PK, $\overline{m}, m$.*
4. *$\mathscr{S}^*$ is given oracle access to user $\mathscr{U}$ to engage in the blind signature protocol with it, mediated by the challenger.*
5. *Let σ be the signature returned by $\mathscr{U}$. σ is hidden from $\mathscr{S}^*$*
6. *$\mathscr{S}^*$ outputs $m' \in \{0,1\}^\lambda$.*

Adversary $\mathscr{S}^*$ wins this game whenever $m' = m$.

The signature scheme is message hiding *when every possible adversary $\mathscr{S}^*$ wins this game with at most negligible probability (i.e., probability at most $2^{-\lambda}$), where the probability is computed over the coin-flips of $\mathscr{S}^*$ and the private coin-flips of $\mathscr{U}$.*

Theorem 3.6. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is message indistinguishable according to definition 3.2. Then Σ is message hiding according to defintion 3.4.*

Proof. Suppose not. So there is an adversarial signer $\mathscr{S}^*$ for the game defined in definition 3.4. We turn it into an adversarial signer $\mathscr{S}^{**}$ for the game defined in definition 3.2 as follows.

1. $\mathscr{S}^{**}$ starts $\mathscr{S}^*$, which returns PK and $\overline{m}$.
2. $\mathscr{S}^{**}$ essentially operates as the challenger for $\mathscr{S}^*$ using whatever it learns in the process to solve its own challenge.
3. $\mathscr{S}^{**}$ does the following. It generates two fresh private messages m_0, m_1 and uses the public message $\overline{m}$ it got from S^* and forwards these together with PK received from $\mathscr{S}^*$ to its own challenger in definition 3.2. This challenger sets up a user with input PK, $\overline{m}_1, m_b$ (depending on its hidden coin flip b) to which $\mathscr{S}^{**}$ is given oracle access to, to engage in the blind signature protocol. $\mathscr{S}^{**}$ forwards this oracle access to $\mathscr{S}^*$.
4. After $\mathscr{S}^*$ has finished interacting with its oracles, is outputs a guess m' (to S^{**}). When $m' = m_{b'}$ as in step 3 for $b' \in 0,1$, $\mathscr{S}^{**}$ returns b' otherwise it returns a random bit.

If $\mathscr{S}^*$ guesses m' correctly, then $m' = m_b$ given to user $\mathscr{U}$ as part of $\mathscr{S}^{**}$ challenge in step 3. The probability that this happens is non-negligible. We conclude that the advantage of $\mathscr{S}^{**}$ guessing b is also non-negligible.

3.4 Signature unlinkability

We now turn to the definition of signature unlinkability. The challenge is to define it in such a way that it does not immediately imply the message indistinguishability property (and thus would be almost equivalent to the general blindness property). We solve this by letting the challenger generate the messages to be signed and giving the signer only the resulting signatures in random order.

Definition 3.5 (Signature unlinkability). *Consider a signature scheme $\langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ and the following game between an adversarial signer $\mathscr{S}^*$ and two honest users $\mathscr{U}_0$ and $\mathscr{U}_1$.*

1. *Run $\mathscr{G}(1^\lambda)$ to generate sk and PK. Give sk, PK to $\mathscr{S}^*$.*
2. *Adversary $\mathscr{S}^*$ outputs PK, and a public message $\overline{m}$, and gives them to the challenger.*
3. *The challenger generates two messages*[10] *m_0, m_1 and sets up user $\mathscr{U}_0$ with input $(PK, \overline{m}, m_0)$ and user $\mathscr{U}_1$ with input $(PK, \overline{m}, m_1)$.*
4. *$\mathscr{S}^*$ is given oracle access to both users to engage in the blind signature protocol with both of them, mediated by the challenger.*
5. *Let σ_0 be the result returned by $\mathscr{U}_0$ and σ_1 be the result returned by $\mathscr{U}_1$.*
6. *If any of the signatures is invalid, the challenger gives $\perp$ to $\mathscr{S}^*$.*[11] *Otherwise the challenger randomly selects $b \in \{0,1\}$ and sets $\bar{b} = 1 - b$. The challenger gives σ_b and $\sigma_{\bar{b}}$ to $\mathscr{S}^*$ in that order.*
7. *$\mathscr{S}^*$ outputs $b' \in \{0,1\}$.*

Adversary $\mathscr{S}^*$ wins this game whenever $b' = b$. The signature scheme is signature unlinkable *when every possible adversary $\mathscr{S}^*$ wins this game with at most negligible advantage (i.e., probability $1/2 \pm 2^{-\lambda}$), where the probability is computed over the coin-flips of $\mathscr{S}^*$ and the private coin-flips of $\mathscr{U}_0$ and $\mathscr{U}_1$.*

We note that Chaum's untraceable payment scheme [6] uses a blind signature scheme that is strongly message hiding and is signature unlinkable as well.

The following signature unlinkable signature scheme (which is a slight modification of Chaum's blind signature scheme) is useful in the proofs of some of the following theorems. We omit the public message $\overline{m}$ for simplicity.

[10] Again not necessarily distinct.

[11] We cannot allow one of these outputs to be $\perp$ instead of a real signature, as that would trivially allow an adversary to distinguish the instance used to generate it. See the discussion in the footnote for definition 3.1

Construction 3.7 (Signature unlinkable signature scheme) *Define a signature unlinkable signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ as follows. Let h_1, h_2 be two cryptographic hash functions.*

$\mathscr{G}$ generates a RSA key pair, and publishes the public key (n, e) while giving the corresponding private key (d, n) to the signer.

A user submitting m for signing first computes $r = h_1(m)$ and then sends $m' = h_2(m)r^e \bmod n$ to the signer. The signer computes $\sigma' = m'^d \bmod n$ and returns it to the user. The user computes $\sigma = \sigma'/r$ as the final signature. This defines $\mathscr{S}$ and $\mathscr{U}$.

$\mathscr{V}$ takes as input m and σ and returns whether $\sigma^e \bmod n = h_2(m)$.

The careful observer will have noted that this is essentially Chaum's blind signature protocol with r derived from m (making it no longer blind as we shall see shortly) while m cannot be recovered from the signature by hiding it using h_2.

Lemma 3.3. *The signature scheme from construction 3.7 is signature unlinkable according to definition 3.5.*

Proof. The construction matches the (syntactic) constraints of definition 2.1 (disregarding the public message $\overline{m}$), and it is easily seen to be complete as defined in 2.2 using the fact that we have $(r^e)^d \bmod n = 1$ in RSA, and the result $\sigma = h_2(m)^d \bmod n$ is a traditional RSA signature over $h(m')$.

This signature scheme is signature unlinkable. As in the game defined in definition 3.5 the challenger generates m_0 and m_1, the adversarial signer $\mathscr{S}^*$ does not know them. By playing the game $\mathscr{S}^*$ learns:

- $m'_0 = h_2(m_0)r_0^e \bmod n$ (and that it is computed by $\mathscr{U}_0$),
- $m'_1 = h_2(m_1)r_1^e \bmod n$ (and that it is computed by $\mathscr{U}_1$),
- $\sigma_0 = h_2(m_0)^d \bmod n$ and $\sigma_1 = h_2(m_1)^d \bmod n$ given in the order defined by a random bit b.

$\mathscr{S}^*$ needs to guess b based on this information (and its knowledge of the public key (n, e)).

As h_1 and h_2 are random oracles, the value $\mathscr{S}^*$ learns for m'_0 could actually correspond to $h_2(m_1)r_1^e \bmod n$ (and vice versa). So the information it relies on to decide on the value for b could just as well be used to argue for the opposite value.

We first show that blindness implies signature unlinkability.

Theorem 3.8. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is blind according to definition 3.1. Then Σ is signature unlinkable according to definition 3.5.*

Proof. Suppose not. So there is an adversarial signer $\mathscr{S}^*$ for the game defined in definition 3.5. We turn it into an adversarial signer $\mathscr{S}^{**}$ for the game defined in definition 3.1 as follows.

1. $\mathscr{S}^{**}$ starts $\mathscr{S}^*$, which returns PK and $\overline{m}$.
2. $\mathscr{S}^{**}$ generates two distinct messages m_0, m_1 and sends them to the challenger along with PK and $\overline{m}$.
3. The challenger randomly selects $b \in \{0,1\}$ and sets $\bar{b} = 1-b$. It sets up user $\mathscr{U}_0$ with input $(\mathrm{PK}, \overline{m}, m_b)$ and user $\mathscr{U}_1$ with input $(\mathrm{PK}, \overline{m}, m_{\bar{b}})$.
4. $\mathscr{S}^{**}$ engages in the blind signature protocol with both users, mediated by the challenger. It does so by relaying all messages to and from $\mathscr{S}^*$.[12]
5. Let σ_b be the result returned by $\mathscr{U}_0$ and $\sigma_{\bar{b}}$ be the result returned by $\mathscr{U}_1$. If both signatures are valid, then the challenger gives (σ_0, σ_1) to $\mathscr{S}^{**}$ in that order by definition.[13] Otherwise it returns $\perp$ to $\mathscr{S}^{**}$.
6. $\mathscr{S}^{**}$ forwards σ_0 and σ_1 in that order to $\mathscr{S}^*$ as the challenge.
7. $\mathscr{S}^*$ outputs $b' \in \{0,1\}$, which $\mathscr{S}^{**}$ forwards as its own output for this challenge.

We observe that if $\mathscr{S}^*$ outputs b' it believes the first signature (σ_0, corresponding to m_0) given as a challenge was generated while interacting with user $\mathscr{U}_{b'}$. Which is the case if b' equals b generated by the challenger for the game defined in definition 3.1. This means that b' is also the correct response to the challenge given to $\mathscr{S}^{**}$. This shows that the advantage of $\mathscr{S}^{**}$ is the same of that of $\mathscr{S}^*$, i.e., non-negligible, contradicting the premise of the theorem.

The signature scheme from construction 3.7 allows us to prove that the converse does not hold: there are signature unlinkable signature schemes that are not blind as the following theorem demonstrates. This shows that also signature unlinkability (like message indistinguishability) is a strictly weaker notion.

Theorem 3.9. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is signature unlinkable according to definition 3.5. This does* not *imply that Σ is blind according to definition 3.1.*

[12] Observe how we use the fact that in definition 3.5 the challenger in the signature unlinkability game (here simulated by $\mathscr{S}^{**}$) randomly selects m_0 and m_1 before giving them to $\mathscr{U}_0$ and $\mathscr{U}_1$ without revealing them to $\mathscr{S}^*$: if $\mathscr{S}^{**}$'s own challenger picks $b = 1$, it is as if S^{**} (simulating $\mathscr{S}^*$'s challenger) picked m_1 (for $\mathscr{U}_0$) and m_0 (for $\mathscr{U}_1$) instead.

[13] Note how $\mathscr{S}^{**}$'s own challenger therefore swaps the signatures generated by the users based on its hidden random bit b.

Proof. Consider the signature scheme from construction 3.7, which is signature unlinkable according to lemma 3.3.

This scheme is clearly not blind: using its knowledge of m_0 (that the adversary chooses according to definition 3.1) the adversarial signer $\mathscr{S}^*$ can compute $r_0 = h_1(m_0)$ and hence $m'_0 = h_2(m_0)r_0^e \bmod n$ that either user $\mathscr{U}_0$ or user $\mathscr{U}_1$ will submit for signing. This allows $\mathscr{S}^*$ to tell which of two users was given m_0 as input by the challenger, and therefore allows $\mathscr{S}^*$ to correctly guess b.

A very similar proof can be used to prove the following theorem.

Theorem 3.10. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is signature unlinkable according to definition 3.5. This does* not *imply that Σ is message indistinguishable according to definition 3.2.*

Proof. Again consider the signature scheme from construction 3.7, which is signature unlinkable according to lemma 3.3.

This signature scheme is not message indistinguishable according to definition 3.2 however. In the message indistinguishability game the adversarial signer knows m_0 and m_1 and therefore can compute $r_b = h_1(m_b)$ and $m'_b = h_2(m_b)r_b^e \bmod n$ for $b \in \{0,1\}$. It can therefore tell which of the two messages the challenger submits for signing and hence can always correctly guess b and win the game.

The reverse is also true.

Theorem 3.11. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is message indistinguishable according to definition 3.2. This does* not *imply that Σ is signature unlinkable according to definition 3.5.*

Proof. Consider the message indistinguishable signature scheme in construction 3.1, where the message to be signed is first CCM encrypted as $c = \{m\}_{k_{\mathscr{U}}}$ under a random key $k_{\mathscr{U}}$. The signer creates the intermediate signature $\sigma' = [c \,\|\, \overline{m}]_{k_{\mathscr{S}}}$. It returns this to $\mathscr{U}$ who adds $k_{\mathscr{U}}$ to create the final signature $\sigma = (\sigma', k_{\mathscr{U}})$.

This scheme is however not signature unlinkable according to definition 3.5. Suppose the adversarial signer keeps the intermediate signatures σ'_0 and σ'_1 it generated while interacting with user $\mathscr{U}_0$ and user $\mathscr{U}_1$ respectively. As in the proof of theorem 3.3 it can match these with $\sigma_b = (\sigma'_b, k_{\mathscr{U}_b})$ and $\sigma_{\bar{b}} = (\sigma'_{\bar{b}}, k_{\mathscr{U}_{\bar{b}}})$ and hence guess b correctly.

This shows that message indistinguishability and signature unlinkability are indeed separate notions.

We will now explore the relationship between signature unlinkability and other notions defined in this paper. For example, what is the relationship between signature unlinkability and message hiding? The blind signature scheme underlying the Idemix attribute based credential scheme [8, 5] is in fact only strongly message hiding but not signature unlinkable.[14] This proves the following theorem.

Theorem 3.12. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is strongly message hiding according to definition 3.3. This does* not *imply that Σ is signature unlinkable according to definition 3.5.*

The other way around, signature unlinkability does imply (weak) message hiding (which explains why we need the slightly stronger notion of message indistinguishability).

Theorem 3.13. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is signature unlinkable according to definition 3.5. Then Σ is message hiding according to definition 3.4.*

Proof. The proof is very similar to the proof of theorem 3.6.

Suppose not. So there is an adversarial signer $\mathscr{S}^*$ for the game defined in definition 3.4. We turn it into an adversarial signer $\mathscr{S}^{**}$ for the game defined in definition 3.5 as follows.

1. $\mathscr{S}^{**}$ starts $\mathscr{S}^*$, which returns PK and $\overline{m}$.
2. $\mathscr{S}^{**}$ essentially operates as the challenger for $\mathscr{S}^*$ using whatever it learns in the process to solve its own challenge.
3. $\mathscr{S}^{**}$ forwards $\overline{m}$ to its own challenger. This challenger generates two messages m_0 and m_1 and sets up a user $\mathscr{U}_0$ with input $(\mathrm{PK}, \overline{m}_i, m_0)$ and a user $\mathscr{U}_1$ with input $(\mathrm{PK}, \overline{m}, m_1)$. $\mathscr{S}^{**}$ is given oracle access to both users to engage in the blind signature protocol with both of them, mediated by the challenger. For $\mathscr{U}_0$ it forwards oracle access to $\mathscr{S}^*$. For $\mathscr{U}_1$, $\mathscr{S}^{**}$

[14] The CL signature over a message $(m_0, \ldots, m_k)$ equals (A, e, v) such that

$$A = \left(Z \prod_{i=0}^{k} R_i^{m_i} S^v \right)^{1/e} \bmod n \; .$$

where Z, R_i, S and n are part of the public key. A and e are generated by the signer, which makes this scheme trivially signature linkable. When submitting a message for signing, the user submits a commitment $\prod_{i=0}^{k} R_i^{m_i} S^{v'}$ to that message to hide it. And to use such a credential in an unlinkable fashion, the main goal of Idemix, the user does not reveal A and e but simply proves their existence to the verifier in zero knowledge.

interacts with this oracle itself. This way $\mathscr{S}^*$ is set up exactly as in the definition of the game in 3.4

4. After $\mathscr{S}^*$ has finished interacting with its oracles, it outputs a guess m' (to S^{**}).
   For the signature unlinkability game S^{**} is playing, S^{**} asks for its challenge. If both signatures (σ_0 generated by $\mathscr{U}_0$ and σ_1 generated by $\mathscr{U}_1$) in step 3 are valid it receives σ_b and $\sigma_{\bar{b}}$ (depending on the private coin flip b of its challenger) in that order.
   It then checks whether σ_b or $\sigma_{\bar{b}}$ is a valid signature over m' (the guess returned by $\mathscr{S}^*$). In the first case it returns $b' = 0$, in the second case it returns $b' = 1$. If neither is the case it returns a random bit b'.

By assumption with some non-negligible probability, m' returned by $\mathscr{S}^*$ corresponds to the oracle set up by $\mathscr{S}^{**}$ in step 3. Then $m' = m_0$ (as $\mathscr{S}^*$ never interacted with $\mathscr{U}_1$). So if m' matches σ_b (the first signature in its challenge), σ_b must be a signature over m_0 and hence $b = 0$. And if it matches $\sigma_{\bar{b}}$, then $b = 1$ instead. We see that in this case $b' = b$ and hence the adversary wins. As we already concluded that this case happens with non-negligible probability, the conclusion follows.

The reverse of this theorem does not hold, by 3.11 and 3.6.

3.5 Message indistinguishability and signature unlinkability

We have so far shown that signature blindness can be separated into two separate properties, message indistinguishability and signature unlinkability, that are indeed independent: one does not imply the other, and neither on its own implies blindness. The natural question to ask is whether message indistinguishability and signature unlinkability together do imply blindness. That would be a nice conclusion, as it would show that the proposed separation is ideal in the sense that both properties capture all what makes a signature scheme blind. Unfortunately, this is not the case if we do not rule out pathological cases of misbehaving users, as the following theorem shows.

Theorem 3.14. *Consider a signature scheme $\Sigma = \langle \mathscr{G}, \mathscr{S}, \mathscr{U}, \mathscr{V} \rangle$ that is message indistinguishable according to definition 3.2 and signature unlinkable according to definition 3.5. This does* not *(in general) imply that Σ is blind according to definition 3.1.*

Proof. Let Σ be a blind signature scheme according to 3.1. Modify Σ as follows to create a new signature scheme Σ'. Pick a particular message $\tilde{m}$.

If σ is the signature returned by user $\mathscr{U}$ when interacting with $\mathscr{S}$, define $\mathscr{U}'$ to return the tuple (σ, β) where β is a random identity *except* when $\mathscr{U}'$ wants a signature on message $\widetilde{m}$. In that case β equals the identity of $\mathscr{U}$.

Clearly, Σ' is no longer blind. The adversarial signer can always commit to messages $m_0 = \widetilde{m}$ and m_1 to the challenger. Depending on its private bit b, the challenger gives $\widetilde{m}$ to either $\mathscr{U}_0$ or $\mathscr{U}_1$. Whichever it is, it will return a signature (σ_0, b) over $\widetilde{m}$ while the other returns (σ_1, β) over m_1 where β is random.

When challenged, the adversary receives $(\sigma_0, b), (\sigma_1, \beta)$ in that order. It returns the b it finds in the first signature which by construction is always equal to the private bit chosen by the challenger. In other words, the adversary wins.

In the message indistinguishability game of definition 3.2, the adversary doesn't receive the final signatures. Therefore its view when interacting with $\mathscr{S}'$ is exactly the same as when interacting with $\mathscr{S}$. We conclude that Σ' is also message indistinguishable.

In the signature unlinkability game of definition 3.5, the adversary does not get to pick the messages to be signed. Instead, the challenger does. With overwhelming probability, $\widetilde{m}$ is not among the messages chosen by the adversary. As a result, the β component of both challenge signatures is random and can be ignored, i.e. the advantage of the adversary against Σ' is no better than against Σ. We conclude that Σ' is signature unlinkable.

4 Conclusions

A summary of our results is presented in figure 1, where we write $A \rightarrow B$ when A implies B. And we write $A \vdash B$ when $B \not\rightarrow A$, i.e., if A does not logically follow from B (or, B does not imply A).

Compiling this figure, we made use of the following transitivity rules governing the relationships among the several notions we defined in this paper.

- $A \rightarrow B$ and $B \rightarrow C$ implies $A \rightarrow C$.
- $A \vdash B$ and $B \rightarrow C$ implies $A \vdash C$.
- $A \rightarrow B$ and $B \vdash C$ implies $A \vdash C$.

As can be seen from the picture, this paper shows that signature blindness can be decomposed into two separate and indeed independent properties: message indistinguishability and signature unlinkability. The more natural notion of message hiding cannot be used for this purpose as it is implied by signature unlinkability.

Unfortunately combining signature unlinkability and message indistinguishability does not give back blindness, although this appears to be the case only in pathological cases. We have so far been unable to prove a restricted version of such a theorem ruling out certain classes of users, and neither did we find a less pathological counterexample. This is left for further research.

I am grateful to the anonymous reviewers for their comments and suggestions that really helped improve the paper.

5 Data deposition statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] M. Abe and E. Fujisaki. “How to date blind signatures”. In: *ASIACRYPT 96* (Kyongju, Korea, Nov. 3–7, 1996). LNCS 1163. Springer, 1996, pp. 244–251.

[2] M. Abe and T. Okamoto. “Provably Secure Partially Blind Signatures”. In: *CRYPTO 2000* (Santa Barbara, CA, USA, Aug. 20–24, 2000). LNCS 1880. Springer, 2000, pp. 271–286.

[3] M. Bellare and C. Namprempre. “Authenticated Encryption: Relations among notions and analysis of the generic composition paradigm”. In: *Cryptology ePrint Archive* 2000/025 (2000). Revised 2007-06-14, p. 25.

[4] S. Brands. *Rethinking Public Key Infrastructures and Digital Certificates; Building in Privacy*. 1st ed. ISBN 0-262-02491-8. MIT Press, 2000.

[5] J. Camenisch and A. Lysyanskaya. “An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation”. In: *EUROCRYPT 2001* (Innsbruck, Austria, May 5–10, 2001). LNCS 2045. Springer, 2001, pp. 93–118.

[6] D. Chaum. “Blind Signatures for Untraceable Payments”. In: *CRYPTO ’82* (Santa Barbara, CA, USA, Aug. 23–25, 1982). Plenum Press, New York, 1982, pp. 199–203.

[7] J.-H. Hoepman. “Privacy Friendly E-Ticketing For Public Transport”. CoRR abs/2101.09085. Jan. 2021. arXiv: 2101.09085 [cs.CR].

[8] IBM Research Zürich Team. *Specification of the Identity Mixer Cryptographic Library*. Report. Version 2.3.4. Zürich: IBM Research, Feb. 2012.

[9] J. Jonsson. “On the Security of CTR + CBC-MAC”. In: *SAC 2002* (St. John’s, Newfoundland, Canada, Aug. 15–16, 2002). LNCS 2595. Springer, 2006, pp. 76–93.

[10] A. Juels, M. Luby, and R. Ostrovsky. “Security of Blind Digital Signatures (Extended Abstract)”. In: *CRYPTO ’97* (Santa Barbara, CA, USA, Aug. 17–21, 1997). LNCS 1294. Springer, 1997, pp. 150–164.

[11] J. Katz and Y. Lindell. *Introduction to Modern Cryptography, Second Edition*. Boca Raton: CRC Press, 2014.

[12] T. Okamoto. “Efficient Blind and Partially Blind Signatures Without Random Oracles”. In: *TCC 2006* (New York, NY, USA, Mar. 4–7, 2006). LNCS 3876. Springer, 2006, pp. 80–99.

[13] T. Okamoto. “Efficient Blind and Partially Blind Signatures Without Random Oracles”. In: *Cryptology ePrint Archive* 2006/102 (2006), p. 34.

[14] T. P. Pedersen. “Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing”. In: *CRYPTO ’91* (Santa Barbara, CA, USA, Aug. 11–15, 1991). LNCS 576. Springer, 1991, pp. 129–140.

[15] D. Schröder and D. Unruh. “Security of Blind Signatures Revisited”. In: *Public Key Cryptography - PKC 2012* (Darmstadt, Germany, May 21–23, 2012). LNCS 7293. Springer, 2012, pp. 662–679.

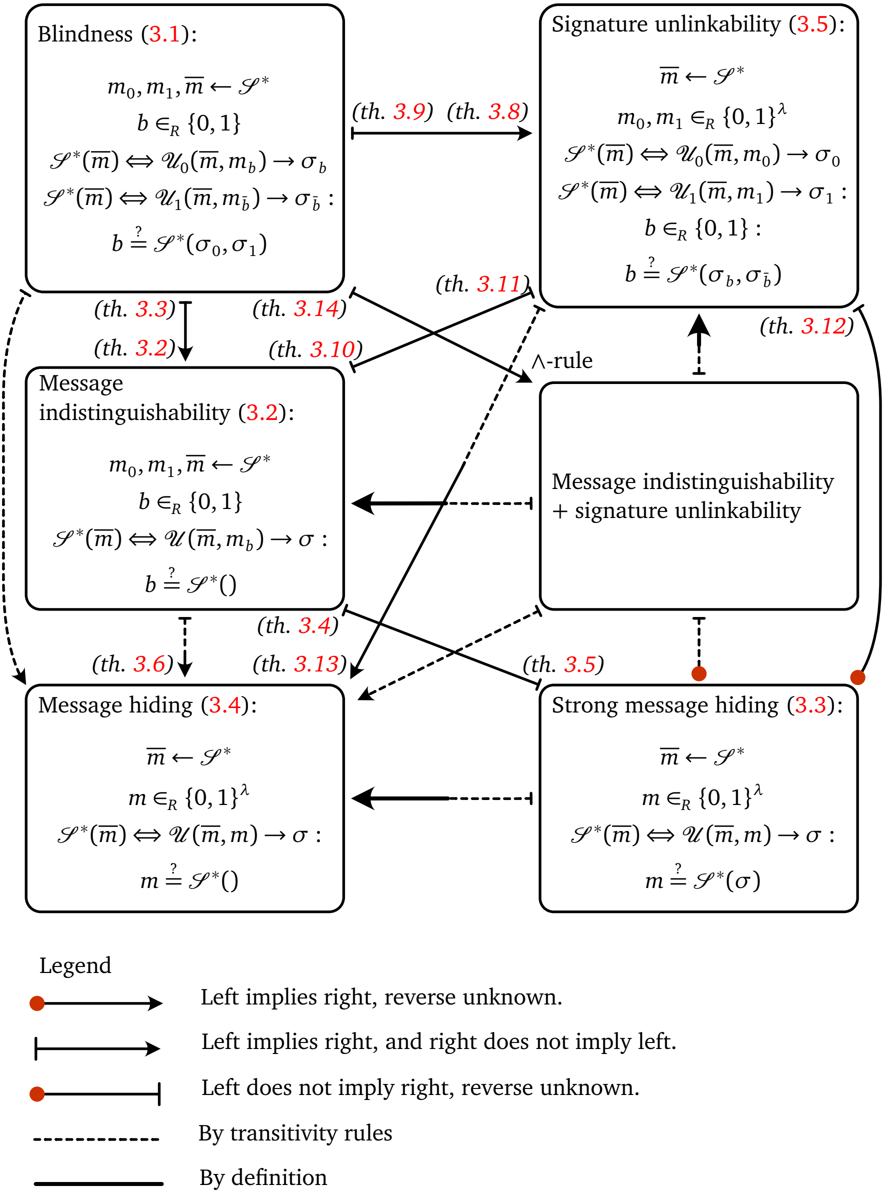
Blindness (3.1):
$m_0, m_1, \overline{m} \leftarrow \mathscr{S}^*$
$b \in_R \{0,1\}$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}_0(\overline{m}, m_b) \rightarrow \sigma_b$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}_1(\overline{m}, m_{\bar{b}}) \rightarrow \sigma_{\bar{b}}$:
$b \stackrel{?}{=} \mathscr{S}^*(\sigma_0, \sigma_1)$
Signature unlinkability (3.5):
$\overline{m} \leftarrow \mathscr{S}^*$
$m_0, m_1 \in_R \{0,1\}^\lambda$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}_0(\overline{m}, m_0) \rightarrow \sigma_0$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}_1(\overline{m}, m_1) \rightarrow \sigma_1$:
$b \in_R \{0,1\}$:
$b \stackrel{?}{=} \mathscr{S}^*(\sigma_b, \sigma_{\bar{b}})$
(th. 3.9) (th. 3.8)
(th. 3.11)
(th. 3.3)
(th. 3.14)
(th. 3.2)
(th. 3.10)
(th. 3.12)
∧-rule
Message indistinguishability (3.2):
$m_0, m_1, \overline{m} \leftarrow \mathscr{S}^*$
$b \in_R \{0,1\}$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}(\overline{m}, m_b) \rightarrow \sigma$:
$b \stackrel{?}{=} \mathscr{S}^*()$
Message indistinguishability + signature unlinkability
(th. 3.4)
(th. 3.6)
(th. 3.13)
(th. 3.5)
Message hiding (3.4):
$\overline{m} \leftarrow \mathscr{S}^*$
$m \in_R \{0,1\}^\lambda$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}(\overline{m}, m) \rightarrow \sigma$:
$m \stackrel{?}{=} \mathscr{S}^*()$
Strong message hiding (3.3):
$\overline{m} \leftarrow \mathscr{S}^*$
$m \in_R \{0,1\}^\lambda$
$\mathscr{S}^*(\overline{m}) \Leftrightarrow \mathscr{U}(\overline{m}, m) \rightarrow \sigma$:
$m \stackrel{?}{=} \mathscr{S}^*(\sigma)$
Legend
Left implies right, reverse unknown.
Left implies right, and right does not imply left.
Left does not imply right, reverse unknown.
By transitivity rules
By definition

Fig. 1. Summary of relations