# Relations Among Notions of Non-Malleability for Encryption

Rafael Pass[1], abhi shelat[2], and Vinod Vaikuntanathan[3]

[1] Cornell
[2] U. Virginia
[3] MIT

**Abstract.** Since its introduction in the early 90's, the notion of non-malleability for encryption schemes has been formalized using a number of conceptually different definitional approaches—most notably, the "pragmatic" indistinguishability-based approach and the "semantical" simulation-based approach. We provide a full characterization of these approaches and consider their robustness under composition.

**Keywords:** Public-key Encryption, Non-malleability.

## 1 Introduction

The basic goal of an encryption scheme is to guarantee the *privacy* of data. A good formalization of privacy is the notion of *semantic security* as defined by Goldwasser and Micali [GM84]. Intuitively, semantic security guarantees that "whatever a polynomial-time machine can learn about a message given its encryption, it can learn even without seeing the encryption."

When encryption schemes are deployed in more complex environments, the demands for security of encryption grow beyond just the basic privacy requirement. Motivated by practical security requirements, the seminal work of Dolev, Dwork and Naor [DDN00] defined the notion of *non-malleability*—a qualitatively stronger notion of security for encryption schemes. In addition to the normal "privacy" guarantee, non-malleability ensures that it is infeasible for an adversary to *modify* a vector of ciphertexts $\alpha_1, \ldots, \alpha_n$ into other ciphertexts of messages which are related to the decryption of $\alpha_1, \ldots, \alpha_n$. This stronger notion of security is critical for many practical applications.

*Two Formalizations.* The notion of non-malleability for encryption schemes has been formalized using two different approaches:

– **The "Semantical" Simulation-based Approach**. The definition presented in the original work of [DDN00] is a so-called "simulation-based" one. The main idea is to capture the requirement that an adversary having access to ciphertexts (and potentially a decryption oracle in case of CCA1/CCA2 attacks), will not be able to "cause more harm" than a simple adversary that does not see any ciphertexts and does not have access to a decryption oracle. This simulation-based definition of non-malleability is denoted SIM-NME, and like semantic security, the goal of this definition is to capture the "meaning" of non-malleability. As a result, it is often harder to directly prove that a scheme meets the simulation-based definition.

– **The "Pragmatic" Indistinguishability-based Approach**. Bellare et.al. present a "comparison-based" formalization of non-malleability [BDPR98]. This notion does away with the "simulator" used in [DDN00] and instead captures non-malleability through an indistinguishability-style definition. Other indistinguishability-based definitions appear in [BS99,PSV06]. We denote by IND-NME the indistinguishability-based approach to defining non-malleability. The goal of this indistinguishability-based approach is to provide definitions that are easier to "work with."

Just as Goldwasser and Micali [GM84] show equivalence between simulation-based and indistinguishability-based definitions of secrecy, Bellare and Sahai [BS06] (clarifying [BS99]) show an equivalence between the simulation-based and the indistinguishability-based approach to defining non-malleability. As we discuss later, their proof however makes certain implicit assumptions on the type of encryption schemes used. As far as we know, equivalences for general encryption schemes are not known.

*Composition and Invalid Ciphertexts.* In practice, encryption schemes must guarantee security also when an adversary receives encryptions of *multiple* messages. It is well known that for the traditional definition of secrecy, "single-message" security implies "multi-message" security – we say that the definition is *closed under composition*. It would be desirable to have a definition of non-malleability that composes (i.e., for which non-malleability for a single message implies non-malleability for multiple messages).

It turns out that this property is highly sensitive to the way non-malleability is formalized. As pointed out by Pass, shelat and Vaikuntanathan [PSV06], there is some ambiguity in the original work of Dolev, Dwork and Naor [DDN00] about how to treat an adversary that sometimes produces *invalid* ciphertexts as part of its output. Whereas the intuitive description of the "spirit" of non-malleability considers an adversary successful if it is able to output ciphertexts that are related to the ciphertexts it receives, the formal definition does not consider an adversary who outputs an invalid ciphertext (even if this event is correlated with the input ciphertexts it receives). It is shown in [PSV06] that for the case of *chosen-plaintext* attacks, this (seemingly minor) issue becomes critical in certain (traditional) applications, and is also essential for proving composability of non-malleability. In both situations a stronger definition, which does not automatically fail an adversary which outputs an invalid ciphertext, is sufficient, whereas the weaker (traditional one) is not. We denote by SIM-NME$'$, IND-NME$'$ these stronger variants of SIM-NME, IND-NME (which are in-line with the definitions of [PSV06,BS06]).

### 1.1 Our results

We may thus broadly categorize definitions of non-malleability into two major groups: "simulation-based" and "indistinguishability-based," and each with two sub-groups: "invalid-allowing" and "invalid-prohibiting." In this paper we first fully characterize the relationship among the different definitional approaches outlined above. Secondly, we consider the robustness of each of the definitions under a natural (and highly desirable) notion of composition. Our motivation is to clarify the definitional imbroglio surrounding the notions. To so do, we present a unified way of defining non-malleability according the above-mentioned different approaches. We furthermore believe that our definitions provide the simplest and cleanest way to formalize non-malleability according to these approaches.

**Relations Between Definitions.** Our results are as follows.

1. *The Case of Invalid-Allowing Definitions* For the case of invalid-allowing definitions, we obtain a separation between the simulation-based definition of non-malleability, $\mathsf{SIM\text{-}NME}'$, and indistinguishability-based definition, $\mathsf{IND\text{-}NME}'$. In particular, under CCA1 or CCA2 attacks, $\mathsf{SIM\text{-}NME}'$ is *strictly* stronger than $\mathsf{IND\text{-}NME}'$, whereas under CPA attacks they are equivalent.
2. *The Case of Invalid-Prohibiting Definitions* For the case of invalid-prohibiting definitions, the simulation-based definition, $\mathsf{SIM\text{-}NME}$ is *equivalent* to the indistinguishability-based definition $\mathsf{IND\text{-}NME}$, under all attacks (i.e., CPA, CCA1 and CCA2).
3. *The Relation between Invalid-Allowing and -Prohibiting Definitions* The first approach to defining non-malleability is *strictly* stronger than the second one. In fact, this holds under all attacks in the simulation-based notion, and under $\mathsf{CCA1}$ and $\mathsf{CPA}$ attacks for the indistinguishability-based notion.

A full characterization of the different definitions is summarized in the table below. The starred results appear in either [DDN00] and/or [BDPR98].

| ATTACK | RELATIONSHIPS |
|---|---|
| CCA2 | $\mathsf{SIM\text{-}NME}' > \mathsf{IND\text{-}NME}' = \mathsf{SIM\text{-}NME} =^* \mathsf{IND\text{-}NME} =^* \mathsf{IND}$ |
| CCA1 | $\mathsf{SIM\text{-}NME}' > \mathsf{IND\text{-}NME}' > \mathsf{SIM\text{-}NME} = \mathsf{IND\text{-}NME} >^* \mathsf{IND}$ |
| CPA | $\mathsf{SIM\text{-}NME}' = \mathsf{IND\text{-}NME}' > \mathsf{SIM\text{-}NME} = \mathsf{IND\text{-}NME} >^* \mathsf{IND}$ |

*Results Concerning Practical Schemes and Restricted Message Spaces.* Many practical and efficient encryption schemes only work for *restricted message spaces*. For example, the El Gamal and Cramer-Shoup schemes work for messages that are elements of some finite group. While it seems natural for the above equivalences to also hold for this special class of encryption schemes, we show in §5 that this intuition is not true. In particular, we show that also for the case of CCA2 attacks, $\mathsf{SIM\text{-}NME}$ is strictly stronger than $\mathsf{IND\text{-}NME}$. Thus, somewhat surprisingly, *For restricted message spaces, "simple"* IND-CCA2 *security does not imply the original semantical (simulation-based) definition of non-malleability.*

This stands in sharp contrast to the result of [DDN00,BDPR98] showing that IND-CCA2 indeed is equivalent SIM-NME for the case of full messages spaces.

*Why Simulation-based Non-malleability is Desirable.* Many practical system attacks such as buffer overflows rely on creating a situation in which a process is fed unexpected input. With this in mind, consider an encryption scheme which has been dutifully designed so that an adversary cannot produce a ciphertexts which decrypt to a certain output value (say $\perp$). A system designer might employ this scheme in a process, and rely on the fact that such inputs cannot be produced by the decrypting algorithm for the correctness of the process.

Now suppose that the adversary might have a way to implement a CCA2 attack. A cryptographer may be content to prove that their encryption scheme is IND-CCA2-secure. However, the systems' practitioner may require something more. She would like the guarantee that even if the adversary has a decryption oracle, the adversary will be unable to "do any more harm" than if the adversary did not have the decryption oracle. In other words, the adversary will be unable to produce unexpected outputs in this case as well—and so the practitioner's original assumptions are still valid. In essence, the situation calls for simulation-based security.

*Remarks.* As shown by Canetti [Can01], a Universally Composable (UC) implementation of an "idealized" encryption functionality $\mathcal{F}_{pke}$ is equivalent to IND-CCA2-secure encryption. Furthermore, the UC definition of security is a semantical notion which provides security under arbitrary concurrent executions; in particular UC security provides security with respect to man-in-the-middle attacks. However, the definition of $\mathcal{F}_{pke}$ allows a corrupted sender to make an honest receiver decrypt a ciphertext to any arbitrary string (and not only those in the domain of the encryption function) *even if this was not possible in a stand-alone setting*; as such UC encryption does not satisfy the above desiderata. We also mention that Goldreich [Gol04] presents a similar semantical (simulation-based) definition of non-malleability, which is equivalent to (simple) indistinguishability under CCA2 attacks; this definition too does not prevent a corrupted sender from making an honest receiver decrypt a ciphertext to any arbitrary string.[4]

*Additional equivalences.* To further clarify the semantical relation between the various notions, we present additional equivalences for certain restricted encryption schemes: Concisely, a scheme which is IND-NME secure and for which it is possible to efficiently produce a ciphertext which decrypts to every output in

[4] On a high-level, the difference between SIM-NME and the definition of [Gol04] is that in the latter, the simulator is required to output *plaintexts* that are indistinguishable from the messages the adversary encrypts, whereas in the former the simulator must do the same as the adversary and output ciphertexts.

the range of the decryption function is also (multi-message) SIM-NME′ secure.[5] Thus, for encryption schemes satisfying certain technical conditions all the above notions are equivalent. In light of this our separation results might seem "artificial".[6] Note, however, that although these restriction are not implausible, they are far from being satisfied all "practical" encryption schemes. Indeed, whereas RSA-OAEP satisfies them (at the cost of "truncating" the message space), other schemes such as CS1 from [CS98] does not.

**Composability of Definitions.** The table below summarizes new and known results regarding the composability of of the various definitional approaches. A $\surd$-mark indicates that the definitions composes, **X**-mark indicates it does not, and ? indicates that the status is unknown. Pass, shelat, and Vaikuntanathan [PSV06] show the $^*$ result. Gennaro and Lindell [GL03] show the $^\dagger$ result. All other results are new in this paper. These new results show that, contrary to folklore belief, *indistinguishability*-based definitions of encryption do *not* necessarily compose in the context of non-malleability.

| | SIM-NME′ | IND-NME′ | SIM-NME | IND-NME |
|---|---|---|---|---|
| CCA2 | ? | $\surd$ | $\surd$ | $\surd$ |
| CCA1 | ? | $\surd$ | **X** | **X** |
| CPA | $\surd$ | $\surd^*$ | $\mathbf{X}^\dagger$ | **X** |

**Related Work.** The work of [BS06] (clarifying the original work of [BS99]) provided a comprehensive study of equivalence between indistinguishability-based and simulation-based definitions. Their main results show such an equivalence for the case of *invalid-allowing* definitions. We here note that their result implicitly makes the assumption that the encryption schemes considered have the property that it is "easy" (i.e., there is a prescribed polynomial-time algorithm) to generate invalid ciphertexts. In contrast, we consider general encryption schemes (i.e., without any such restriction). Interestingly, we show that the notions no longer are equivalent when doing so (furthermore, when considering restricted messages spaces, equivalence does not hold even if there exists a prescribed polynomial-time algorithm for generating invalid ciphertexts).

Nevertheless, we emphasize that proof techniques from [BS99,BS06] are useful also when considering general encryption schemes. Indeed, our equivalence proof for the case of *invalid-prohibiting* definitions (i.e., showing that SIM-NME = IND-NME) borrows from their original proof.[7]

[5] This result generalizes the earlier results by [BS06]. See Section 1.1 for more details.

[6] In a sense all separation results can be called either "artificial" or "trivial"—if they are satisfied by known schemes then they are trivial, otherwise they are "artificial".

[7] The original published version of their results [BS99] claimed an equivalence between SIM-NME and an indistinguishability-based definition of non-malleability due to [BDPR98]. This claim was later retracted in the new version [BS06] (due to subtleties pointed out by Lindell). We mention, however, that our definition of IND-NME is (seemingly) different from the indistinguishability-based definition of [BDPR98].

We also mention that various other definitions of non-malleability for encryption schemes have been proposed (e.g [BDPR98,BS06,Gol04]). Our goal is not to fully characterize the relative strength of all variants of non-malleability. Rather, we highlight the differences between certain natural definitional approaches (i.e., simulation v.s. indistinguishability, and invalid-allowing v.s. invalid-prohibiting).

## 2 Definitions

*Oracles.* In a chosen-plaintext attack ($\mathsf{CPA}$), the oracles $O_1, O_2$ return the empty string. In a $\mathsf{CCA1}$ attack, the oracle $O_1(\text{PK}, \cdot)$ returns the decryption of its input under public key PK (which is implicit by context). Finally, in a $\mathsf{CCA2}$ attack, both oracles return decryptions with the exception that $O_2(\text{PK}, \boldsymbol{y}, \cdot)$ returns $\perp$ when queried on a ciphertext contained in $\boldsymbol{y}$.

*Comparing Definitions.* If $\text{D1}, \text{D2}$ are two definitions, the notation $\text{D1} > \text{D2}$ means that: "Every scheme $\Pi$ which satisfies $D1$ also satisfies $D2$, and if there exists a scheme $\Pi$ which satisfies D2, then there exists a scheme $\Pi'$ which also satisfies D2 but does not satisfy D1." We say that $\text{D1} = \text{D2}$ if the set of schemes that satisfy D1 is identical to the set of schemes that satisfy D2.

### 2.1 Simulation-based Definitions of Non-Malleable Encryption

**Definition 1 ($\mathsf{SIM\text{-}NME}'$ Security).** *Define the following two experiments.*

$$\begin{array}{ll}
\mathsf{SIM\text{-}NME}'(\Pi, A, k, \ell, r) & \overline{\mathsf{SIM\text{-}NME}'}(\Pi, S, k, \ell, r) \\
\quad (\text{PK}, \text{SK}) \leftarrow \mathsf{Gen}(1^k) & \quad (\text{PK}, \text{SK}) \leftarrow \mathsf{Gen}(1^k) \\
\quad (M, s) \leftarrow A_1^{O_1}(\text{PK}) & \quad \boxed{(M, s) \leftarrow S_1(\text{PK})} \\
\quad (m_1, \ldots, m_\ell) \stackrel{\$}{\leftarrow} M(1^k) & \quad (m_1, \ldots, m_\ell) \stackrel{\$}{\leftarrow} M(1^k) \\
\quad \boldsymbol{y} \leftarrow \mathsf{Enc}(\text{PK}, \boldsymbol{m}) & \\
\quad ((c_1, \ldots, c_r), \sigma) \leftarrow A_2^{O_2}(\boldsymbol{y}, h(\boldsymbol{m}), s) & \quad \boxed{((c_1, \ldots, c_r), \sigma) \leftarrow S_2(h(\boldsymbol{m}), s)} \\
\quad d_i = \begin{cases} \text{COPY} & \textit{if } c_i \in \boldsymbol{y} \\ \mathsf{Dec}(\text{SK}, c_i) & \textit{o.w.} \end{cases} & \quad d_i = \begin{cases} \text{COPY} & \textit{if } c_i = \text{COPY} \\ \mathsf{Dec}(\text{SK}, c_i) & \textit{o.w.} \end{cases} \\
\quad \textit{Output } (M, \boldsymbol{m}, (d_1, \ldots, d_r), \sigma) & \quad \textit{Output } (M, \boldsymbol{m}, (d_1, \ldots, d_r), \sigma)
\end{array}$$

*Here $M$ is a Turing machine that samples a vector of $\ell(k)$ messages from a distribution. We say that $M$ is an $(p, \ell)$-valid message-sampler if 1) the running-time of $M(1^k)$ is bounded by $p(k)$, and 2) there exists polynomials $l_1, l_2, .., l_\ell$ such that $M(1^k)$ always outputs message sequences $(m_1, \ldots, m_{\ell(k)})$ such that $|m_i| = l_i(1^k)$ for all $1 \leq i \leq \ell(k)$.*

*An encryption scheme $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is $\mathsf{SIM\text{-}NME}'$-secure if for polynomials $\ell(k)$, $r(k)$ and $p(k)$, every polynomial-time computable history function $h(\cdot)$, every p.p.t. adversary $A = (A_1, A_2)$ which runs in time $p(k)$ and always outputs a $(p, \ell)$-valid message sampler, there exists a p.p.t. algorithm $S = (S_1, S_2)$*

*that always outputs a* $(p,\ell)$*-valid message sampler, such that the following two distributions are computationally indistinguishable:*

$$\left\{\mathsf{SIM\text{-}NME}'(\Pi, A, k, \ell(k), r(k))\right\}_k \stackrel{c}{\approx} \left\{\overline{\mathsf{SIM\text{-}NME}'}(\Pi, S, k, \ell(k), r(k))\right\}_k \quad (1)$$

We also define a weaker notion of this definition named SIM-NME by requiring that the outputs of the two experiments are indistinguishable only for a certain "restricted" set of adversaries $A$. Define the following two types of adversaries:

1. *non-copying adversaries*: $A = (A_1, A_2)$ is said to be non-copying if in the above experiment $A_2$ never outputs a ciphertext $c_i$, s.t., $c_i \in \boldsymbol{y}$.
2. *valid adversaries*[8]: $A$ is said to be valid if in the above experiment $A$ only outputs ciphertexts that are in the range of the encryption function (on input PK), i.e., it holds that for all $c_i$, there exists an $d_i$ such that $c_i \in \mathsf{Enc}(\text{PK}, d_i)$.

**Definition 2 (SIM-NME Security).** *An encryption scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is* SIM-NME*-secure if for polynomials* $\ell(k)$*,* $r(k)$ *and* $p(k)$*, every polynomial-time computable history function* $h(\cdot)$*, every* **non-copying, valid** *p.p.t. adversary* $A = (A_1, A_2)$ *which runs in time* $p(k)$ *and always outputs a* $(p,\ell)$*-valid message sampler, there exists a p.p.t. algorithm* $S = (S_1, S_2)$ *that always outputs a* $(p,\ell)$*-valid message sampler, such that the ensembles in equation (1) are indistinguishable to any p.p.t. distinguisher* $D$*.*

*Single-message Versus Many-message Security.* We have presented definitions which allow the adversary to see a sequence of encrypted messages. Forboth the above definitions of non-malleability, a scheme satisfying the definition in the case when $\ell(k) = 1$ (but $r(k)$ is still arbitrary), is said to be *single-message* secure. The question of whether any single-message secure scheme is also (many-message) secure is the question of composability of the definition.

*Remarks.* Single-message SIM-NME security is a rewriting of the original DDN simulation-based definition of non-malleability. The main difference between our definition and definition of DDN is that we dispense with the relation $R$ and instead use the notion of indistinguishability of the outputs. This difference is inconsequential (since any p.p.t distinguisher can be described as a p.p.t relation and vice versa); however, this draws a parallel to the (upcoming) indistinguishability-based definition of non-malleability, which we term IND-NME. In this way, we emphasize the meaning of this definition: neither a ciphertext of a chosen message or a decryption oracle can substantially alter an adversaries ultimate "behavior." Given this interpretation, it is also intuitive to see why the valid-adversary is somehow artificial. Moreover this restriction prevents the definition from composing—i.e., it is possible for a scheme to be single-message SIM-NME secure, but not SIM-NME secure. We also remark that our definition of single-message SIM-NME$'$ security is syntactically equivalent to the SNM definition of non-malleability from [BS06].

[8] This interpretation comes from [DDN00] where they write "$\mathcal{A}$ produces...ciphertexts $(f_1, \ldots)$...with $f_i \in \mathsf{Enc}(\beta_i)$..."

### 2.2 Indistinguishablility-based Definitions

The following definition of non-malleability was introduced in [PSV06] and is syntactically very close to the definition of [BS99,BS06].

**Definition 3 ($\mathsf{IND\text{-}NME}'$ Security [PSV06]).** *Let $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme and let the random variable $\mathsf{IND\text{-}NME}_b(\Pi, A, k, \ell, r)$ where $b \in \{0,1\}$, $A = (A_1, A_2)$ and $k, \ell, r \in \mathbb{N}$ denote the result of the following probabilistic experiment:*

$\mathsf{IND\text{-}NME}'_b(\Pi, A, k, \ell, r)$ :
$(\text{PK}, \text{SK}) \leftarrow \mathsf{Gen}(1^k)$
$((m_{0,1}, \ldots, m_{0,\ell}), (m_{1,1}, \ldots, m_{1,\ell}), s) \leftarrow A_1^{O_1}(\text{PK})$ *s.t.* $|m_{0,i}| = |m_{1,i}|$
$y_i \leftarrow \mathsf{Enc}(\text{PK}, m_{b,i})$ *for* $i \in [1, \ell]$
$(c_1, \ldots, c_r) \leftarrow A_2^{O_2}(\boldsymbol{y}, s)$
*Output* $(d_1, \ldots, d_r)$ *where* $d_i = \begin{cases} \text{COPY} & \text{if } c_i \in \boldsymbol{y} \\ \mathsf{Dec}(\text{SK}, c_i) & \text{otherwise} \end{cases}$

$(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is $\mathsf{IND\text{-}NME}'$-secure if $\forall$ p.p.t. algorithms $A = (A_1, A_2)$ and for any polynomials $\ell(k)$ and $r(k)$, the following two ensembles are computationally indistinguishable:*

$$\left\{\mathsf{IND\text{-}NME}'_0(\Pi, A, k, \ell(k), r(k))\right\}_k \overset{c}{\approx} \left\{\mathsf{IND\text{-}NME}'_1(\Pi, A, k, \ell(k), r(k))\right\}_k \quad (2)$$

We also introduce a weaker version of this definition, $\mathsf{IND\text{-}NME}$, in which, as in the previous section, (2) need only hold for non-copying, valid adversaries $A$.

**Definition 4 ($\mathsf{IND\text{-}NME}$ Security).** *An encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is $\mathsf{IND\text{-}NME}$-secure if $\forall$* **non-copying, valid** *p.p.t. algorithms $A = (A_1, A_2)$ and for any polynomials $\ell(k)$ and $r(k)$, the ensembles in the equation (2) are computationally indistinguishable.*

*Single-message Security.* For both the above indistinguishability-based definitions, we obtain the weaker notion of *single-message* security by restriction attention to the case when $\ell(k) = 1$. We also note that our definition of single-message $\mathsf{IND\text{-}NME}'$ security is a syntactical rewriting of (and thus equivalent to) the definition of IND-PAX of [BS06].

## 3 Equivalences Between Definitions

**Theorem 1.** $\mathsf{SIM\text{-}NME} = \mathsf{IND\text{-}NME}$ *for all attacks.*

The equivalence proof for this theorem uses ideas from Bellare and Sahai [BS99]. Note however that it does not show that $\mathsf{SIM\text{-}NME}' = \mathsf{IND\text{-}NME}'$ (as was the goal in Bellare and Sahai's revised paper [BS06]). Let us briefly recall the subtle issue in the original proof in [BS99] (the same issue appears in the revised proof that $\mathsf{SIM\text{-}NME}' = \mathsf{IND\text{-}NME}'$ in [BS06]). In one step of the equivalence proof, the $\mathsf{SIM\text{-}NME}$ simulator must re-encrypt a vector of ciphertexts which the adversary has produced. If an "aborting" adversary has produced an invalid ciphertext, it is not clear whether the simulator can proceed—in particular, the encryption scheme $\Pi$ *might not provide* an efficient method available to produce an invalid ciphertext (as was the case in the previous section). The proof does hold, however, for a valid adversary who always produces ciphertexts that are in the range of the Enc function.

In the full version, we present a direct equivalence proof for SIM-NME and IND-NME which is simple and extends to the case of many-message security. Moreover, the proof also leads to the following corollary relating SIM-NME′ and IND-NME′ used in Theorem 4:

**Corollary 1.** *If $\Pi$ is* SIM-NME′*-secure, then $\Pi$ is also* IND-NME′ *secure.*

For completeness, we present a proof of the following theorem in the full version which has been partially shown by Dolev, Dwork, and Naor [DDN00].

**Theorem 2.** IND-NME′-CCA2 = IND-NME-CCA2 = IND-CCA2.

In the weaker CPA attack, we show that the simulation and indistinguishability definitions for invalid ciphertext-producing adversaries are also equivalent by adapting a simpler version of Thm. 1. This implies that the construction from [PSV06] meets the strongest notion of non-malleability for the CPA attack. The proof appears in the full version.

**Theorem 3.** *Under a CPA attack,* SIM-NME′ = IND-NME′.

## 4 Separating the SIM-NME′ and IND-NME′ Definitions

**Theorem 4 (Main Separation).** *Under* CCA1 *or* CCA2 *attacks,* SIM-NME′ > IND-NME′ *even for single-message security.*

Corollary 1 shows that SIM-NME′ implies IND-NME′. Thus, the main idea for this separation is to design an encryption scheme in which the set of messages for which a ciphertext can be efficiently computed and the range of the decryption function *differ*. As one concrete example below, we design an IND-NME′ scheme in which it is nearly impossible for an adversary to produce a ciphertext which decrypts to $\perp$ (i.e., an invalid ciphertext) *unless* it has adaptive access to a decryption oracle. [9] We show the scheme so constructed meets the IND-NME′ definition. However, it does not meet the SIM-NME′ definition under a CCA1 or CCA2 attack, because an adversary (with access to a decryption oracle) is able to produce a ciphertext that decrypts to $\perp$ whereas a simulator (without access to a decryption oracle) is unable. Thus, the outputs of the SIM-NME′ and $\overline{\mathsf{SIM\text{-}NME}'}$ experiments will be trivially distinguishable. The general idea behind these type of arguments first appears in [DDN00] and is also used in [BDPR98] to show other separations.

[9] Another example would be a finite message space, i.e., a message space which includes all strings in $\{0,1\}^k$ and a scheme in which the range of the decryption function includes one $k^2$ bit string. We discuss this later in §5.

*Proof.* Let $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme that satisfies $\mathsf{IND\text{-}NME'}$ under a $\mathsf{CCA}$ attack. Consider encryption scheme $\Pi'$ defined in the figure below. The key property of $\Pi'$ is that $\mathsf{Dec}'$ never outputs $\perp$ unless it is queried with

> ENCRYPTION SCHEME $\Pi'$
>
> $\mathsf{Gen}'(1^k)$ : Run $(\text{PK}, \text{SK}) \leftarrow \mathsf{Gen}(1^k)$. Pick random $k$-bit string $\alpha$ and set $\text{SK}' \leftarrow (\text{SK}, \alpha)$.
>
> $\mathsf{Enc}'(\text{PK}', m)$ : Run $c \leftarrow \mathsf{Enc}(\text{PK}, m)$. Output $(1, 0^k, c)$ as ciphertext.
>
> $\mathsf{Dec}'(\text{SK}', c')$ : Parse $c'$ as $(b, \beta, c)$ where $b$ is a bit, $\beta$ is a $k$-bit string.
>
> 1. If $b = 0$ and $\beta = 1^k$, then output $\alpha$.
> 2. If $b = 0$ and $\beta = \alpha$, then output $\perp$.
> 3. If $b = 1$ and $\beta = 0^k$, run $m \leftarrow \mathsf{Dec}(\text{SK}, c)$. If the output is $\perp$, output 0. Otherwise, output $m$.
> 4. Otherwise, output 0.

the special "open sesame" string $\alpha$, and a decryption oracle is necessary to learn the "open sesame" string.

It is easy to see that $\Pi'$ syntactically is an encryption scheme. The only issue is to argue that $\Pi'$ is perfectly correct, which follows because perfect correctness only applies to decryption of honestly encrypted messages (which are never invalid ciphertexts).

*Claim.* $\Pi' = (\mathsf{Gen}', \mathsf{Enc}', \mathsf{Dec}')$ meets the $\mathsf{IND\text{-}NME'\text{-}CCA}$ definition.

*Proof.* Suppose there exists an adversary $A'$ which breaks the $\mathsf{IND\text{-}NME'\text{-}CCA}$ definition for $\Pi'$. Such an adversary can be used to construct an adversary $A$ which breaks the $\mathsf{IND\text{-}NME'\text{-}CCA}$ definition for $\Pi$ as follows:

The new adversary $A$ simulates $(\mathsf{Gen}', \mathsf{Enc}', \mathsf{Dec}')$ for $A'$ by picking $\alpha$ itself and using the oracles for $\mathsf{Dec}$ to answer queries. More precisely, on input a public key PK, $A$ generates a $k$-bit string $\alpha$ and feeds PK to $A'$. When $A'$ asks decryption queries, $A$ simulates the $\mathsf{Dec}'$ algorithm by using $\alpha$ as the second component of $\text{SK}'$ and the decryption oracle in order to compute $\mathsf{Dec}(c, \text{SK})$. When $A'$ produces two challenge messages, $A$ forwards these messages along, and when it receives a challenge ciphertext $y$, $A$ feeds $(1, 0^k, y)$ to $A'$. In the case of a $\mathsf{CCA2}$ attack, $A$ again simulates the $\mathsf{Dec}'$ function, and when $A'$ finally returns an answer, $A$ echoes it. $A$ perfectly simulates the $\mathsf{IND\text{-}NME'\text{-}CCA}$ game for $A'$, and thus succeeds with exactly the same probability as $A'$.

*Claim.* $\Pi'$ does not meet the $\mathsf{SIM\text{-}NME'\text{-}CCA}$ definition.

*Proof.* Consider the relation $R(x, \boldsymbol{x}, M, s)$ which is 1 if $x$ is $\perp$ and 0 otherwise.

A $\mathsf{CCA1}$ adversary with access to a decryption oracle can satisfy $R$ by making a decryption query on the message $(0, 1^k, 0)$ to get the value $\alpha$, and then by outputting the ciphertext $(0, \alpha, 0)$.

However, it is not possible for a simulator $S$ without access to the decryption oracle to satisfy $R$. Such a simulator only has an exponentially small chance of guessing the correct $\alpha$ string necessary to produce $\perp$. Thus, $\Pi'$ will not satisfy $\mathsf{SIM\text{-}NME'\text{-}CCA1}$.

### 4.1 More Separations for CCA1 and CPA Attacks

We now show that $\mathsf{IND\text{-}NME'}$ is stronger than $\mathsf{IND\text{-}NME}$ *when considering weaker* $\mathsf{CPA}$ *and* $\mathsf{CCA1}$ *attacks*. Recall that $\mathsf{IND\text{-}NME'}$ and $\mathsf{IND\text{-}NME}$ are different only in that the former protects against all PPT adversaries, whereas the latter protects against only valid PPT adversaries.[10] By combining the equivalence from Theorem 1, we also get a separation between $\mathsf{IND\text{-}NME'}$ and $\mathsf{SIM\text{-}NME}$. For $\mathsf{CCA2}$ attacks, they become equivalent (See Thm. 2).

**Theorem 5.** $\mathsf{IND\text{-}NME'} > \mathsf{IND\text{-}NME}$ *for CCA1 and CPA attacks even for single-message security.*

**Corollary 2.** $\mathsf{IND\text{-}NME'} > \mathsf{SIM\text{-}NME}$ *for CCA1 and CPA attacks even for single-message security.*

*Proof.* (Of Corollary 2.) By Theorem 5, $\mathsf{IND\text{-}NME'} > \mathsf{IND\text{-}NME}$ for CCA1 and CPA attacks and by Theorem 1, $\mathsf{SIM\text{-}NME} = \mathsf{IND\text{-}NME}$ under all attacks.

The main idea for the proof of Theorem 5 is to use the *DDN-lite* transformation [Dwo99,Nao04] to transform an $\mathsf{IND\text{-}NME}$-secure encryption scheme into one that remains $\mathsf{IND\text{-}NME}$-secure (Claim 4.1), but is vulnerable to an $\mathsf{IND\text{-}NME'}$ attack (Claim 4.1).

We actually prove a stronger statement which gives us a way to transform an $\mathsf{IND\text{-}CPA}$-secure encryption scheme into one that is $\mathsf{IND\text{-}NME}$-secure. While this result has been claimed in [Dwo99,Nao04], as far as the authors know, a proof has never been printed. Our proof also shows that the construction also transforms an $\mathsf{IND\text{-}CCA1}$ scheme into an $\mathsf{IND\text{-}NME\text{-}CCA1}$ scheme. The $\mathsf{IND\text{-}NME'}$-attack against this scheme is an adaptation of the attack against DDN-lite, given in [PSV06].

*Proof.* (of Theorem 5) Let $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme that is $\mathsf{IND\text{-}CPA}$-secure (respectively, $\mathsf{IND\text{-}CCA1}$-secure). Let $\Sigma = (\mathsf{Gen}_{sig}, \mathsf{Sign}, \mathsf{Ver})$ be a strongly unforgeable one-time signature scheme. Such a signature scheme can be constructed from one-way functions (The existence of one-way functions, in turn, is implied by the existence of a $\mathsf{IND\text{-}CPA}$-secure encryption scheme). We construct a new encryption scheme $\Pi_{\mathrm{L}}$ from $\Pi$ and show that $\Pi_{\mathrm{L}}$ satisfies the $\mathsf{IND\text{-}NME}$ definition but does does not satisfy $\mathsf{IND\text{-}NME'}$.

*Claim.* $\Pi_{\mathrm{L}}$ meets the $\mathsf{IND\text{-}NME}$ definition.

[10] We say that an invalid ciphertext "decrypts" to $\perp$ (Bot) and hence the title of the subsection.

ENCRYPTION SCHEME $\Pi_{\mathrm{L}}$

$\mathsf{Gen}'(1^k)$ : Run $\mathsf{Gen}(1^k)$ $2k$ times with independent random coins to produce $2k$ pairs of keys $(\mathrm{PK}_b^i, \mathrm{SK}_b^i)$ for $i \in [1,k]$ and $b \in \{0,1\}$. Let $\mathrm{PK}' = \left[\mathrm{PK}_b^i\right]_{i\in[k],b\in\{0,1\}}$ and $\mathrm{SK}' = \left[\mathrm{SK}_b^i\right]_{i\in[k],b\in\{0,1\}}$

$\mathsf{Enc}'(m, \mathrm{PK}')$ : Run $\mathsf{Gen}_{sig}(1^k)$ to generate a key-pair $(\mathrm{VKSIG}, \mathrm{SKSIG})$ for the signature scheme. Let VKSIG a $k$-bit string, and let the $i^{th}$ bit of VKSIG be denoted $\mathrm{VKSIG}_i$.
Run $c_i \leftarrow \mathsf{Enc}(\mathrm{PK}_i^{\mathrm{VKSIG}_i}, m)$ for $i \in [1,k]$.
Let $\sigma \leftarrow \mathsf{Sign}(\mathrm{SKSIG}, (c_1, c_2, \ldots, c_k))$.
Output $\left[(c_1, \ldots, c_k), \mathrm{VKSIG}, \sigma\right]$ as the ciphertext.

$\mathsf{Dec}'(c', \mathrm{SK}')$ : Parse $c'$ as $((c_1, \ldots, c_k), \mathrm{VKSIG}, \sigma)$.
If $\mathsf{Ver}(\mathrm{VKSIG}, (c_1, \ldots, c_k), \sigma) = \mathrm{REJECT}$, output $\perp$.
Otherwise, decrypt the $c_i$'s with the corresponding secret-keys to get corresponding messages $m_i$. If all $m_i$'s are equal, output $m_1$, else output $\perp$.

*Proof.* First, we show that an encryption scheme $\widetilde{\Pi}$, constructed from $\Pi$ in the following way, meets the IND-CPA (respectively, IND-CCA1) definition (Proposition 1). $\widetilde{\Pi} = (\widetilde{\mathsf{Gen}}, \widetilde{\mathsf{Enc}}, \widetilde{\mathsf{Dec}})$ is constructed as follows:

1. $\widetilde{\mathsf{Gen}}$ runs $k$ copies of $\mathsf{Gen}$ to generate public-keys $\widetilde{\mathrm{PK}} = (\mathrm{PK}_1, \mathrm{PK}_2, \ldots, \mathrm{PK}_k)$ and corresponding secret-keys $\widetilde{\mathrm{SK}} = (\mathrm{SK}_1, \mathrm{SK}_2, \ldots, \mathrm{SK}_k)$.
2. $\widetilde{\mathsf{Enc}}(m, \widetilde{\mathrm{PK}})$ runs $\mathsf{Enc}(m, \mathrm{PK}_i)$ for all $i \in [k]$, with independently chosen randomness, and outputs the vector of $k$ encryptions $\left[c_1, c_2, \ldots, c_k\right]$.
3. $\widetilde{\mathsf{Dec}}(c, \widetilde{\mathrm{SK}})$ parses $c$ as $\left[c_1, c_2, \ldots, c_k\right]$. Let $m_i = \mathsf{Dec}(c_i, \mathrm{SK}_i)$. If all the $m_i$ are the same, output $m_1$. Otherwise, output $\perp$.

Secondly, in Proposition 2, we show that if $\widetilde{\Pi}$ is IND-CPA-secure (respectively, IND-CCA1-secure), then $\Pi'$ is IND-NME-CPA-secure (resp., IND-NME-CCA1-secure). This proof appears in the full version.

**Proposition 1.** *If $\Pi$ is* IND-CPA*-secure (or* IND-CCA1*-secure), then so is $\widetilde{\Pi}$.*

*Proof.* The proof is a straightforward hybrid argument. The only complication stems from the simulation of the oracle in the CCA1 case. When the adversary asks to decrypt a ciphertext $\hat{c} = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_k)$, decrypt $c_j$ using the secret-key $\mathrm{SK}_j$ (if $j \neq i$) and using the decryption oracle for $\mathrm{PK}_i$ (if $j = i$).

**Proposition 2.** *If $\widetilde{\Pi}$ is* IND-CPA*-secure (respectively,* IND-CCA1*-secure), then $\Pi_L$ is* IND-NME*-secure (respectively,* IND-NME-CCA1*-secure).*

*Claim.* $\Pi_{\mathrm{L}}$ is not IND-NME$'$-secure under CPA and CCA1 single-message attacks.

*Proof.* We specify an adversary $A = (A_1, A_2)$ and a distinguisher $D$ such that $D$ distinguishes between $\{\mathsf{IND\text{-}NME}'_0(\Pi, A, k, 1)\}$ and $\{\mathsf{IND\text{-}NME}'_1(\Pi, A, k, 1)\}$. $A$ works as follows:

1. $A_1$ outputs two arbitrary messages $(m_0, m_1)$ and no state information.

2. On input ciphertext $c = \big[(e_1, \ldots, e_k), \text{VKSIG}, \sigma\big]$, let $\text{VKSIG} := b_1 b_2 \ldots b_k$. $A_2$ produces a new ciphertext $c'$ as follows: $A_2$ generates a new signing key $(\text{SKSIG}', \text{VKSIG}')$. Let $\text{VKSIG}' := b'_1 b'_2 \ldots b'_k$. $A_2$ outputs ciphertext $c' = ((x_1, \ldots, x_k), \text{VKSIG}', \sigma')$ where

$$x_i = \begin{cases} e_i & \text{if } b'_i = b_i \\ E_{\text{PK}_i^{b'_i}}(m_0) & \text{otherwise} \end{cases}$$

and $\sigma'$ is the signature of $(x_1, \ldots, x_k)$ under the signing key $\text{SKSIG}'$.

Notice that $\mathsf{NME}_0(\Pi, A, k, 1) = m_0$ and $\mathsf{NME}_1(\Pi, A, k, 1) = \perp$ which can be easily distinguished by a distinguisher $D$ that outputs 0 on $m_0$ and 1 on $\perp$.

## 5 Additional Separations with Finite Message Spaces

Many encryption schemes such as El Gamal, RSA, Cramer-Shoup, and the league of schemes based on elliptic curves and bilinear maps only process messages from a finite message space such as the elements of some group $G$. In order to capture the security of such systems, Cramer and Shoup [CS98] redefine the encryption primitive to incorporate (a) a key-dependent message space $M_{\text{PK}}$ and (b) a p.p.t. message tester algorithm $\mathcal{M}$ that on input $1^k, \text{PK}, \alpha$, determines whether $\alpha$ is an element of the message-space for the security parameter $1^k$ and the public key PK. The encryption algorithm $\mathsf{Enc} : M_{\text{PK}} \to \{0,1\}^*$ now takes an input message from $M_{\text{PK}}$ and produces general bit strings, and the decryption algorithm maps $\{0,1\}^*$ to $M_{\text{PK}} \cup \perp$. The correctness property is only required to hold over the message space.

In this section, however, we note that if the message space is finite, then the previously proven equivalence relationship between the weaker notions of SIM-NME and IND-NME no longer holds. While the particular counter-example that we use for the separation may seem contrived, this separation has practical significance since it runs against our "intuition" about IND-CCA2 security.

The idea behind this separation is as follows. We construct an encryption scheme whose message space includes three elements, $\{0, 1, \chi\}$ where $\chi$ is related to the public key PK. Moreover, we make it difficult for an adversary to learn $\chi$ unless it has a decryption oracle (notice, the definition for finite message space only requires the message space to be easily decidable, but does not require it to be enumerable.[11]) From this point, the argument is the same. Namely, an adversary with an oracle can produce a (valid) ciphertext decrypting to $\chi$ (therefore it is a valid adversary), whereas the simulator can only produce ciphertexts decrypting to 0 or 1. The subtle difference between this argument and the one from §4 is that in this one, it is not the simulator's inability to produce a ciphertext which decrypts to $\perp$, but rather its inability to *learn a special message*

[11] One could require enumerability of the message space. However, it is unclear such a restriction helps; and it is clear that it needlessly prevents us from using more exotic algebraic structures for encryption.

*in the message space* which provides the separation. In the full message case, there are *no* such special messages since any string can be encrypted. This is the reason that the separation can be extended to valid adversaries.

Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be an $\mathsf{IND\text{-}NME}$-secure encryption scheme for general message spaces, and let $f$ be a one-way permutation.[12]

| Finite message space Encryption Scheme $\Gamma$ |
|---|
| $\mathsf{Gen}'(1^k)$ : Run $\mathsf{Gen}(1^k)$ to generate a key pair $(\text{PK}, \text{SK})$. Pick $k$-bit random string $\alpha$ and compute $\beta = f(\alpha)$. Set $\text{SK}' = (\text{SK}, \alpha)$ and $\text{PK}' = (\text{PK}, \beta)$. The message tester $\mathcal{M}(m)$ works as follows: if $m \in \{0, 1\}$ or if $f(m) = \beta$, then return 1. Otherwise, return 0. (The messages space consists of $\{0, 1, \alpha\}$). |
| $\mathsf{Enc}'(m, \text{PK}')$ : if $\mathcal{M}(m) = 0$, return an error. Otherwise, run $c \leftarrow \mathsf{Enc}(\text{PK}, m)$ and return $(1, c)$. |
| $\mathsf{Dec}'(c', \text{SK}')$ : Parse $c'$ as $(b, c)$, and $\text{SK}'$ as $(\text{SK}, \alpha)$. If $b = 0$ then output $\alpha$. Otherwise, output $m \leftarrow \mathsf{Dec}(\text{SK}, c)$. |

$\mathsf{IND\text{-}NME}$ security of the above finite-message space encryption scheme directly follows from the security of $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$. In order to violate $\mathsf{SIM\text{-}NME}$, the adversary $B$ must be non-aborting. Therefore, the final ciphertext it produces must be in the range of the $\mathsf{Enc}$ function (i.e., of the form $(1, c)$). Combined with the one-wayness of $f$, a simulator not having access to a decryption oracle will not be able to construct a *valid* encryption to the message $\alpha$.

However, a CCA1 attacker can easily do so by first querying $(0, 0)$ to find $\alpha$ (notice that the attacker can *query* the oracle on invalid ciphertexts, but cannot produce them as final output), and then honestly encrypting $\alpha$.

## 6 Special Cases for Equivalence

The separation between $\mathsf{SIM\text{-}NME}'$ and $\mathsf{IND\text{-}NME}'$ hinged on the fact that the set of messages for which one can efficiently compute a ciphertext and the range of the decryption procedure differ. When these two sets are made to coincide, a scheme that is $\mathsf{IND\text{-}NME}'$ secure is also $\mathsf{SIM\text{-}NME}'$-secure. Thus, we provide an easy way to prove that a scheme meets the strongest notion of non-malleability. As a corollary, we get that the main construction of [DDN00] achieves the strongest form of security – that is $\mathsf{SIM\text{-}NME}'$-security against $\mathsf{CCA2}$ attacks.

**Theorem 6.** *Any (finite message-space) encryption scheme $\Pi$ which meets the* $\mathsf{IND\text{-}NME}$ *definition and for which there is an efficient algorithm $F$, which on input $(\text{PK}, d)$ where $d$ is a string in the range of* $\mathsf{Dec}$*, produces a ciphertext $c$ such that $d \leftarrow \mathsf{Dec}(\text{SK}, c)$, also meets the* $\mathsf{SIM\text{-}NME}'$ *definition.*

[12] In fact a one-way function would suffice. We only use a permutation for ease of exposition.

This restriction could easily be added to many schemes by taking the message space to be some set $\{0,1\}^{\ell(k)}$ for all keys generated by $\mathsf{Gen}(1^k)$ (and by making it easy to generate invalid ciphertext). We note that the RSA-OAEP padding scheme does exactly this.

## 7 Composition: Many message security

In [PSV06], the authors show that $\mathsf{IND\text{-}NME}'$ security under $\mathsf{CPA}$ attacks composes. That is, if an encryption scheme is $\mathsf{IND\text{-}NME}'$-secure when the adversary receives one encryption, it will also be $\mathsf{IND\text{-}NME}'$-secure in a situation in which the adversary receives many encryptions.

A natural question is whether the same phenomena occurs under stronger $\mathsf{CCA1}$ and $\mathsf{CCA2}$ attacks. In this section, we answer affirmatively as described in the following theorem.

**Theorem 7.** *A scheme $\Pi$ meets* $\mathsf{IND\text{-}NME}'$ *under attack* ATK *iff it meets single-message* $\mathsf{IND\text{-}NME}'$ *under attack* ATK.

*Proof Sketch:* The forward implication follows directly. For the reverse direction, we present a routine hybrid argument that uses an adversary $(A_1, A_2), D$ with advantage $\epsilon$ to construct a new adversary $(A'_1, A'_2), D$ which breaks the single-message security with advantage $\eta/\ell^2$.

Define a new experiment $\mathsf{IND\text{-}NME}'_{(b_1,\ldots,b_\ell)}(\Pi, A, k, \ell)$ indexed by an $\ell$-bit string $(b_1,\ldots,b_\ell)$ which is the same as $\mathsf{IND\text{-}NME}'_0(\Pi, A, k, \ell)$ except in the fourth line (change is underlined): $y_i \leftarrow \mathsf{Enc}\underline{(\textsc{pk}, m_{b_i,i})}$ for $i \in [1,\ell]$. Define

$$B(i) = (\overbrace{0,\ldots,0}^{l-i}, \overbrace{1,\ldots,1}^{i})$$

and note that $\mathsf{IND\text{-}NME}'_0 = \mathsf{IND\text{-}NME}'_{B(0)}$ and $\mathsf{IND\text{-}NME}'_1 = \mathsf{IND\text{-}NME}'_{B(\ell)}$. Because $D$ distinguishes $\mathsf{IND\text{-}NME}'_0$ from $\mathsf{IND\text{-}NME}'_1$, there exists some $g^* \in [1,\ell]$ such that $D$ distinguishes $\mathsf{IND\text{-}NME}'_{B(g^*)}$ from $\mathsf{IND\text{-}NME}'_{B(g^*+1)}$ with advantage $\eta/\ell$. This suggests the following adversary: $A'^{O}_1(\textsc{pk})$ guesses value $g \in [1,\ell]$ and runs $A_1(\textsc{pk})$—answering any decryption queries by using its own decryption oracle—and waits to receive the two vector of messages $(m_{0,1},\ldots,m_{0,\ell})$ and $(m_{1,1},\ldots,m_{1,\ell})$. Finally, $A'$ outputs $(m_{0,g}, m_{1,g})$ as its challenge pair and outputs state information containing $g$ and $\boldsymbol{m}_0, \boldsymbol{m}_1$.

Adversary $A'^{O}_2(y, \textsc{state}')$, on input an encryption $y$, first executes the replaced line 4 of experiment $\mathsf{IND\text{-}NME}'_{B(g)}$ (described above) with the exception that it uses $y$ for the $(g+1)$th encryption: $y_{g+1} \leftarrow y$. This is possible because it receives the messages vectors $\boldsymbol{m}_0$ and $\boldsymbol{m}_1$ in $\textsc{state}'$.

It then feeds the resulting vector of ciphertexts $\boldsymbol{y}$ to $A_2$ to produce another vector of ciphertexts $(c_1,\ldots,c_\ell)$ and uses this vector as its own output. To answer any oracle query $c$ from $A_2$, $A'_2$ uses the following procedure: If $c = y_j$ for any $j \in [1,\ell]$, then return $\perp$. Otherwise, it uses its own decryption oracle to decrypt $c$ and answers with the returned message.

Notice that $\mathsf{IND\text{-}NME}'_0(A'_1, A'_2)$ and $\mathsf{IND\text{-}NME}'_{B(g^*)}(A_1, A_2)$ are *syntactically* the same, as are $\mathsf{IND\text{-}NME}'_1(A'_1, A'_2)$ and $\mathsf{IND\text{-}NME}'_{B(g^*+1)}(A_1, A_2)$. Because $A'$ guesses $g^*$ correctly with probability $1/\ell$, $D$'s overall advantage in breaking the single-message non-malleability is $\eta/\ell^2$. $\square$

One can see here the importance of removing the "valid adversary" restriction for the hybrid argument to work. This follows because the reduction feeds a hybrid distribution to $A_2$ and, *even if $A_2$ is itself a valid adversary* for the multi-message experiment, $A_2$ may produce invalid ciphertexts when it is fed a *hybrid* distribution. Moreover, these $\perp$ values may form the basis for distinguishability in the hybrid experiment. Thus, one cannot guarantee that valid adversaries for the multi-message experiment can be transformed into valid adversaries for the single-message experiment. The separation in Claim 4.1 exploits this issue.[13]

**SIM-NME and IND-NME Do Not Compose Against CCA1 or CPA Attacks** We now show that (if there exist SIM-NME-secure encryption schemes) there is an encryption scheme $\Pi'$ that is SIM-NME or IND-NME-secure when the adversary is given *one* ciphertext as the challenge, but there is an adversary $A'$ that completely breaks the IND-NME-security of $\Pi'$ when given *polynomially many* ciphertexts as challenge.

The encryption scheme $\Pi'$ is simply the encryption scheme constructed in the proof of Thm. 5 (relying on the DDNLite construction). Thm. 5 showed that $\Pi_{\mathrm{L}}$ is 1-message IND-NME-secure (and therefore 1-message SIM-NME-secure). The many-message attack against $\Pi'$ is a simple *covering attack*. (We mention that Gennaro and Lindell [GL03] pointed out that the DDNLite encryption scheme is not secure under under many messages. Although they did not include a description of the attack, we believe they had a similar attack in mind.)

Recall that an encryption of a message $m$ under $\Pi'$ consists of many encryptions of $m$ with respect to a randomly chosen set of $k$ (out of $2k$) public-keys. Given many (roughly $k \log k$) independent encryptions of $m$, the one can essentially recover an encryption of $m$ under *all* the $2k$ public-keys. This will enable us to construct a completely new encryption of $m$, and thus break $\mathsf{IND\text{-}NME}'$ security.

**Theorem 8.** *Let* $\mathsf{atk} \in \{\mathsf{CPA}, \mathsf{CCA1}\}$. *If there exists an encryption scheme that is* IND-atk *secure, then there exists another encryption scheme* $\Pi'$ *that is* 1-*message* IND-NME-atk-*secure (respectively* SIM-NME-atk-*secure), but is not even* IND-NME-CPA-*secure (respectively,* SIM-NME-CPA-*secure).*

*Proof.* Omitted

[13] This argument also applies to a different interpretation of "valid adversary" in which one forces the single-message experiment to return 0 when invalid ciphertexts are produced. In this case, when $A_2$ produces invalid ciphertexts in the hybrid experiments, the value of both hybrid experiments ($b = 0, 1$) will be 0 and the weaker definition will thus be met *even though* there might still be a distinguisher which could have distinguished the output of $A'_2$.

**SIM-NME and IND-NME Compose Under CCA2 Attacks**

**Theorem 9.** *If an encryption scheme $\Pi$ is 1-message IND-NME-CCA2-secure, then it is many-message IND-NME-CCA2-secure.*

The proof of this theorem follows from Theorem 2, which shows that under CCA2 attacks, IND-NME and SIM-NME definitions coincide with the IND-NME$'$ definition, and Theorem 7 which shows that IND-NME$'$ composes under a many-message attack.

*Acknowledgments.* We would like to thank one of the anonymous Crypto referees for thorough and helpful comments.

## References

[BDPR98] M. Bellare, A. Desai, D. Pointcheval, and P. Rogaway. Relations among notions of security for public-key encryption schemes. In *CRYPTO*, 1998.
[BS99] Mihir Bellare and Amit Sahai. Non-malleable encryption: Equivalence between two notions, and an indistinguishability-based characterization. In *CRYPTO*, pages 519–536, 1999.
[BS06] Mihir Bellare and Amit Sahai. Non-malleable encryption: Equivalence between two notions, and an indistinguishability-based characterization. 2006. http://eprint.iacr.org/2006/228.
[Can01] Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *FOCS*, pages 136–145, 2001.
[CS98] Ronald Cramer and Victor Shoup. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In *CRYPTO*, pages 13–25, 1998.
[DDN00] Danny Dolev, Cynthia Dwork, and Moni Naor. Nonmalleable cryptography. *SIAM J. Comput.*, 30(2):391–437, 2000.
[Dwo99] Cynthia Dwork. The non-malleability lectures. Course notes for Stanford CS 359, 1999. http://theory.stanford.edu/g̃durf/cs359-s99/.
[GL03] Rosario Gennaro and Yehuda Lindell. A framework for password-based authenticated key exchange. In *EUROCRYPT*, pages 524–543, 2003.
[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.
[Gol04] Oded Goldreich. *Foundations of Cryptography, Volume 2.* Cambridge University Press, 2004.
[Nao04] Moni Naor. A taxonomy of encryption scheme security. 2004.
[PSV06] Rafael Pass, Abhi Shelat, and Vinod Vaikuntanathan. Construction of a non-malleable encryption scheme from a any semantically secure one. In *CRYPTO*, pages –, 2006.