Non-Malleable Non-Interactive Zero Knowledge and Adaptive Chosen-Ciphertext Security

AMIT SAHAI*

Abstract

We introduce the notion of *non-malleable* non-interactive zero-knowledge (NIZK) proof systems. We show how to transform any ordinary NIZK proof system into one that has strong non-malleability properties. We then show that the elegant encryption scheme of Naor and Yung [NY] can be made secure against the strongest form of chosen-ciphertext attack by using a non-malleable NIZK proof instead of a standard NIZK proof.

Our encryption scheme is simple to describe and works in the standard cryptographic model under general assumptions. The encryption scheme can be realized assuming the existence of trapdoor permutations.

1 Introduction

Modern cryptography provides us with several fundamental tools, from encryption schemes to zero-knowledge proofs. For each of these tools, we have some intuition about what they "should" achieve. But we must be careful to understand the gap between our intuition and what we can actually achieve. Indeed, a major goal of cryptography is to refine our tools to bring them closer to achieving our intuition, while simultaneously refining our intuitions to be consistent with what is attainable.

In this work, we focus on two basic cryptographic tools: non-interactive zero-knowledge proofs and public-key encryption schemes. We refine our intuition behind non-interactive zero-knowledge (NIZK) proofs by defining the notion of *non-malleable* NIZK, and give constructions that achieve non-malleability. We then use non-malleable NIZK to build a simple public key encryption scheme under general assumptions that achieves the highest level of privacy known to be possible, *i.e.* security against adaptive chosen-ciphertext attack. This considerably simplifies the only previously known encryption scheme achieving this level of security under general assumptions.

Non-Malleable Non-Interactive Zero-Knowledge. Zero-knowledge proofs, introduced by Goldwasser, Micali, and Rackoff [GMR], are fascinating and extremely useful constructs. The intuition behind them is clear from their name: they should be convincing, and yet yield nothing beyond the validity of the assertion being proven. Blum, Feldman, and Micali [BFM] extend this seemingly contradictory notion to the non-interactive setting as well; they define a notion of non-interactive zero-knowledge proofs, which are sent without interaction from the Prover to the Verifier, in a model where all parties share a common random reference string. NIZK proofs have proved themselves of great value, and have been used to achieve chosen-ciphertext security for encryption schemes [NY, DDN] as well as signature schemes secure against chosen-message attack [BG].

For NIZK, the formal requirement of [BFM] (later refined by [FLS]) captures the following requirement: what one can output after seeing an NIZK proof is indistinguishable from what one can output without seeing it, *if the output is examined independent of the actual reference string*. However, the reference string is precisely what is used to build and verify NIZK proofs! Thus, nothing in the formal definition prevents the possibility that seeing one NIZK proof could enable an adversary to prove many other statements it could not have proved otherwise, which is very far from the intuition of "zero-knowledge."[1] To some extent, this is unavoidable: one

* MIT Laboratory for Computer Science, 545 Technology Square, Cambridge, MA 02139, USA. E-Mail: amits@theory.lcs.mit.edu. Supported by a DOD/NDSEG Graduate Fellowship and partially by DARPA grant DABT-96-C-0018.

[1] This is true even for the adaptive zero-knowledge definition of NIZK. We give an example in the next paragraph.

can always duplicate an NIZK proof, and hence prove something that one possibly could not have proved beforehand. But can we hope to demand the following requirement: whatever one can prove after seeing an NIZK proof, one could also have proved before seeing it, *except for the ability to duplicate the proof*? This would come much closer to our intuition of "zero-knowledge." Following the paradigm of [DDN] (who studied, among other topics, similar problems which arise in concurrent executions of *interactive* zero-knowledge proofs), we call this property *non-malleability*[2] for non-interactive zero-knowledge, and it is precisely this property we introduce and examine in this work.

Note that this non-malleability property does not follow from the current definitions of NIZK, as the following simple example demonstrates: Suppose π is a NIZK proof system for a hard language $L \in \mathrm{NP}$. Let L' be the language of pairs of strings in L, *i.e.* $L' = \{(x, y) : x \in L \;\mathtt{and}\; y \in L\}$. Then if we define a new proof system Π that uses a reference string $\Sigma = \sigma_1 \circ \sigma_2$ consisting of the concatenation of two reference strings for π, and has proofs simply consist of pairs of proofs under π that $x \in L$ and $y \in L$ using reference strings σ_1 and σ_2 respectively, it is easy to verify that Π will be a NIZK proof system for L'. However, suppose we see a proof $p = (p_1, p_2)$ that $(x, y) \in L'$ and we do not know how to prove that $y \in L$, but we have a witness to the fact that $x' \in L$. Then we can build a proof p_1' under π that $x' \in L$, and by splicing it with the proof we were given, produce a new proof $p' = (p_1', p_2)$ under Π that $(x', y) \in L'$, which we did not know how to do before seeing p.

Our Results on Non-Malleable NIZK. We formalize the notion of non-malleable NIZK, and give a construction that transforms any ordinary NIZK proof system into a non-malleable NIZK proof system, under the assumption that one-way functions exist. Our basic construction achieves non-malleability only with respect to a single proof, *i.e.* the non-malleability is guaranteed when the adversary only sees a single proof from the outside world. We note however, that this suffices for our application of constructing encryption schemes secure against adaptive chosen-ciphertext attack. We then give another construction that achieves non-malleability with respect to any fixed polynomial number of proofs, where the size of the common random reference string grows with the bound on the number of proofs, but the probability of cheating remains negligible.

[2] We choose this term since the definition deals with the ability to modify (or "maul") an NIZK proof to produce different valid proofs. As we noted earlier, this seems to us a *minimal* requirement one should expect from "zero-knowledge" proofs. Indeed, it is fascinating to ask what still stronger properties one could hope to define and achieve.

CCA-Secure Encryption – Discussion. In the context of encryption, which is perhaps the best studied notion in cryptography, our basic intuition is to think of encryption schemes as providing a "secure envelope," which only the proper addressee can open. This is a very compelling metaphor, and is undoubtedly the inspiration for the design of many cryptographic protocols. But what are the essential properties of a "secure envelope"? The most basic is *passive privacy* – that a passive eavesdropper should not learn any useful information about a message from its encryption. Goldwasser and Micali's notion of *semantic security* [GM] is the accepted formalization of this property, and encryption schemes that achieve this property have been studied extensively. However, we may require stronger privacy properties from encryption schemes: If encryption is to be used as a primitive in higher level protocols, we may need security against *active attacks*, such as a chosen-ciphertext attack (CCA), where the adversary has some access to a decryption mechanism. There are two commonly considered notions of chosen-ciphertext attack. In the strongest proposed notion, known as an "adaptive chosen-ciphertext attack" (denoted $\mathrm{CCA2}$), the adversary is allowed to ask for the decryption of any ciphertext other than the challenge ciphertext. In the weaker form, known as a "lunchtime attack" (denoted $\mathrm{CCA1}$), the adversary has access to the decryption mechanism only prior to receiving the challenge ciphertext which it must decipher. (Formal definitions of security against various kinds of attacks are given in Definition 2.3). Security with respect to the stronger notion ($\mathrm{CCA2}$) implies other desirable properties which we do not have space to discuss, such as non-malleability (e.g. see [DDN, BDPR, BS]), as well. This kind of security is needed if encryption is to be used in general applications, such as exchange of e-mail, where users may unwittingly provide attackers with decryptions of selected ciphertexts. Encryption with this strongest property ($\mathrm{CCA2}$-security) has been proposed as a component in authentication and key exchange protocols [BCK], electronic payment [SET], and deniable authentication protocols [DNS]. For more discussion on the importance of chosen-ciphertext security, see [Sho98].

Prior Work on CCA-Secure Encryption. Much work has been done on achieving chosen-ciphertext security in encryption schemes. Naor and Yung [NY] gave an elegant construction based on general cryptographic assumptions which achieves security against the weaker form of chosen-ciphertext attack ($\mathrm{CCA1}$). Rackoff and Simon [RS] showed how to modify the scheme of Naor and Yung to achieve security against adaptive chosen-ciphertext attack ($\mathrm{CCA2}$), but only in a model with a trusted center assigning certified keys to all parties. More recently, Bellare and Rogaway [BR1, BR] have proposed efficient schemes whose security relies on a random oracle, and Cramer and Shoup [CS] have given an efficient scheme based on the Decisional Diffie-Hellman assumption. Until now, the only known encryption scheme achieving adaptive chosen-ciphertext (CCA2) security based on general assumptions was given by Dolev, Dwork, and Naor [DDN].

Our Results on CCA-Secure Encryption. In this work, we show how to use non-malleable NIZK to modify the original elegant scheme of Naor and Yung and achieve provable security against adaptive chosen-ciphertext attack based only on general assumptions. The scheme of Naor and Yung is very simple: A message is encrypted using two independent semantically-secure encryption functions, and an NIZK proof is provided showing that both ciphertexts are encryptions of the same message. Unfortunately, the NIZK proof alone fails to provide security against *adaptive* chosen-ciphertext attack (CCA2).[3] We show that by simply replacing the NIZK proof with a *non-malleable* NIZK proof, one achieves full security against adaptive chosen-ciphertext attack. In contrast, the only previously known scheme based on general assumptions of [DDN] has a quite involved construction, which exploits an intricate interplay betweeen many encryptions, NIZK proofs, and other components. Our scheme gives a simple framework for building encryption schemes secure against CCA2 from two well-defined basic components, namely semantically-secure encryption schemes and non-malleable NIZK proofs. If efficient implementations of non-malleable NIZK proof systems for the consistency of encryptions were found for some particular semantically-secure encryption schemes, this would yield efficient encryption schemes secure against adaptive chosen-ciphertext attack, as well. Based on the current state of knowledge, the NIZK proof system needed for our scheme can be realized based on any trapdoor permutation. Thus trapdoor permutations suffice for realizing our encryption scheme.

Overview. We will first formalize our notion of non-malleable NIZK, as well as a closely related property called simulation soundness. We then present a construction for achieving non-malleable NIZK, and give a generalization, based on polynomials, of our construction to achieve non-malleability against any fixed polynomial number of proofs. Finally, we present the construction of an encryption scheme secure against adaptive chosen-ciphertext attack, and formally prove its correctness.

[3] This can be seen trivially by considering an NIZK proof system which simply ignores the last bit of any proof. Thus, in an adaptive chosen-ciphertext attack, the adversary can simply flip the last bit of the NIZK proof in the challenge ciphertext and query the decryption oracle to break the scheme.

2 Preliminaries

We use standard notations and conventions for writing probabilistic algorithms and experiments. If A is a probabilistic algorithm, then $A(x_1, x_2, \ldots; r)$ is the result of running A on inputs $x_1, x_2, \ldots$ and coins r. We let $y \leftarrow A(x_1, x_2, \ldots)$ denote the experiment of picking r at random and letting y be $A(x_1, x_2, \ldots; r)$. If S is a finite set then $x \leftarrow S$ is the operation of picking an element uniformly from S. $x := \alpha$ is a simple assignment statement. By a "non-uniform (probabilistic) polynomial-time adversary," we always mean a circuit whose size is polynomial in the security parameter. Sometimes we break up algorithms (such as simulators and adversaries) into multiple stages; in such cases we will use κ or τ to denote state information passed from one stage to another.

We first define efficient non-interactive proof systems, and then give a definition of adaptive single-theorem non-interactive zero-knowledge (as in [FLS]):

Definition 2.1 [NIPS] $\pi = (f, \mathrm{P}, \mathcal{V})$ is an *efficient non-interactive proof system* for a language $L \in \mathrm{NP}$ with witness relation R if f is a polynomial and P and $\mathcal{V}$ are probabilistic polynomial-time machines such that:

(Completeness): For all $x \in L$ and all w such that $R(x, w) = \texttt{true}$, for all strings σ of length $f(|x|)$, we have that $\mathcal{V}(x, \mathrm{P}(x, w, \sigma), \sigma) = \texttt{true}$.

(Soundness): For all adversaries A, if $\sigma \in \{0,1\}^{f(k)}$ is chosen randomly, then the probability that $A(\sigma)$ will output (x, p) such that $x \notin L$ but $\mathcal{V}(x, p, \sigma) = \texttt{true}$ is negligible in k.

Definition 2.2 [NIZK] $\pi = (f, \mathrm{P}, \mathcal{V}, \mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2))$ is an *efficient adaptive single-theorem non-interactive zero-knowledge proof system* for the language L if $(f, \mathrm{P}, \mathcal{V})$ is an efficient non-interactive proof system and $\mathcal{S}_1, \mathcal{S}_2$ are probabilistic polynomial-time machines such that for all non-uniform polynomial-time adversaries $A = (A_1, A_2)$, we have that $\left|\Pr[\,\mathsf{Expt}_A(k) = 1\,] - \Pr[\mathsf{Expt}_A^S(k) = 1]\right|$ is negligible in k, where $\mathsf{Expt}_A(k)$ and $\mathsf{Expt}_A^S(k)$ are:

$\mathsf{Expt}_A(k)$:	$\mathsf{Expt}_A^S(k)$:
$\sigma \leftarrow \{0,1\}^{f(k)}$	$(\sigma, \kappa) \leftarrow \mathcal{S}_1(1^k)$
$(x, w, \tau) \leftarrow A_1(\sigma)$	$(x, w, \tau) \leftarrow A_1(\sigma)$
$p \leftarrow \mathrm{P}(x, w, \sigma)$	$p \leftarrow \mathcal{S}_2(x, \kappa)$
$\texttt{return}\ A_2(p, \tau)$	$\texttt{return}\ A_2(p, \tau)$

We also use the standard definitions for encryption schemes secure against adaptive chosen-ciphertext attack (denoted CCA2) and chosen-plaintext attack (denoted CPA), which can be found for example in [BDPR]. Note that semantic security is equivalent to security against chosen-plaintext attack.

Definition 2.3 [CPA, CCA1, CCA2] Let $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ be an encryption scheme and let $A = (A_1, A_2)$ be an adversary. For ATK $\in \{\mathrm{CPA}, \mathrm{CCA1}, \mathrm{CCA2}\}$ and $k \in \mathsf{N}$, define the *advantage* of A to be:

$$\mathsf{Adv}_A^{\mathrm{ATK}}(k) \stackrel{\mathrm{def}}{=} 2 \cdot \Pr[\mathsf{Expt}_A^{\mathrm{ATK}}(k) = 1] - 1$$

where:

$$\boxed{\begin{array}{l} \mathsf{Expt}_A^{\mathrm{ATK}}(k): \\ \quad (pk, sk) \leftarrow \mathcal{G}(1^k) \\ \quad (m_0, m_1, \tau) \leftarrow A_1^{\mathcal{O}_1}(pk) \\ \quad b \leftarrow \{0,1\} \\ \quad y \leftarrow \mathcal{E}_{pk}(m_b) \\ \quad g \leftarrow A_2^{\mathcal{O}_2}(y, \tau) \\ \quad \texttt{return } 1 \text{ iff } g = b \end{array}}$$

If ATK = CPA	then	$\mathcal{O}_1(\cdot) = \varepsilon$
	and	$\mathcal{O}_2(\cdot) = \varepsilon$
If ATK = CCA1	then	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$
	and	$\mathcal{O}_2(\cdot) = \varepsilon$
If ATK = CCA2	then	$\mathcal{O}_1(\cdot) = \mathcal{D}_{sk}(\cdot)$
	and	$\mathcal{O}_2(\cdot) = \mathcal{D}_{sk}^{(y)}(\cdot)$

Above, $\mathcal{D}_{sk}^{(y)}(\cdot)$ means the oracle that decrypts any ciphertext except y. We insist, above, that A_1 outputs m_0, m_1 with $|m_0| = |m_1|$. We say that $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ is secure against ATK if A being non-uniform polynomial-time implies that $\mathsf{Adv}_A^{\mathrm{ATK}}(\cdot)$ is negligible. ▌

3 Non-Malleable NIZK

In this section, we define non-malleable NIZK and related notions. The notion of non-malleability for NIZK is meant to capture the following requirement: "whatever one can prove after seeing an NIZK proof, one could also have proved without seeing it, except for the ability to duplicate the proof." Put a little more formally, suppose we are given an adversary A that, after seeing a proof p of the statement $x \in L$, is able to produce a proof $p' \neq p$ for some x' satisfying some polynomial-time verifiable property $R(x')$, with probability q. Then we should be able to transform A into another adversary A' that *directly* produces a proof for some x' that satisfies $R(x')$, with probability negligibly different than q.

We can turn this into a formal definition of a non-malleable NIZK proof system, which we give with respect to single proofs. It turns out, however, that we will need to capture a stronger notion of non-malleability, which we call *adaptive* non-malleability, where we allow the adversary to ask for the proof of a theorem of its choosing. Note that this is not possible the usual scenario, since some party must supply a witness for every theorem in order for a proof of it to be produced. Of course if the adversary did this, then this would make the definition trivial, since then the adversary can produce the proof on its own and is not receiving any outside help. Hence, we instead make this definition with respect to simulated proofs, which do not require any witnesses.[4] We present here this stronger definition of non-malleability, and defer the weaker definition of ordinary (non-adaptive) non-malleability (which is implied by the definition below) to the full version of this paper.

Definition 3.1 [Adaptive Non-Malleable NIZK] Let $\Pi = (f_\Pi, \mathcal{P}_\Pi, \mathcal{V}_\Pi, \mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2))$ be an efficient non-interactive single-theorem adaptive zero-knowledge proof system for the language L. We say that Π is an *adaptively non-malleable NIZK proof system* for L if there exists an efficient non-interactive proof system $\pi = (f_\pi, \mathcal{P}_\pi, \mathcal{V}_\pi)$ for the same language L, and a probabilistic polynomial-time oracle machine M such that:

For all non-uniform polynomial-time adversaries $A = (A_1, A_2)$ and for all non-uniform polynomial-time relations R, we have that $\left|\Pr\left[\,\mathsf{Expt}_{A,R,\Pi}(k)\,\right] - \Pr\left[\,\mathsf{Expt}'_{A,\pi}(k)\,\right]\right|$ is negligible in k, where $\mathsf{Expt}_{A,R,\Pi}(k)$ and $\mathsf{Expt}'_{A,R,\pi}(k)$ are:

$$\boxed{\begin{array}{l} \mathsf{Expt}_{A,R,\Pi}(k): \\ \quad (\Sigma, \kappa) \leftarrow \mathcal{S}_1(1^k) \\ \quad (x, \tau) \leftarrow A_1(\Sigma) \\ \quad p \leftarrow \mathcal{S}_2(x, \Sigma, \kappa) \\ \quad (x', p', \mathrm{aux}) \leftarrow A_2(x, p, \Sigma, \tau) \\ \quad \texttt{return true iff} \\ \qquad (\ p' \neq p\ )\ \texttt{and} \\ \qquad (\ \mathcal{V}_\Pi(x', p', \Sigma) = \texttt{true}\ )\ \texttt{and} \\ \qquad (\ R(x', \mathrm{aux}) = \texttt{true}\ ) \\ \hline \mathsf{Expt}'_{A,R,\pi}(k): \\ \quad \sigma \leftarrow \{0,1\}^{f_\pi(k)} \\ \quad (x', p', \mathrm{aux}) \leftarrow M^A(\sigma) \\ \quad \texttt{return true iff} \\ \qquad (\ \mathcal{V}_\pi(x', p', \sigma) = \texttt{true}\ )\ \texttt{and} \\ \qquad (\ R(x', \mathrm{aux}) = \texttt{true}\ ) \end{array}}$$

We also define another notion for NIZK which we call simulation soundness, which is similar to but incom-

[4] It is conceivable that we could introduce an all-powerful party that supplies witnesses to true statements, and make the definition this way. As done in [FLS] when defining adaptive zero-knowledge, we choose not to take this route, as it would necessarily give the adversary the power to check membership in L, which is a power we do not necessarily want to capture. We note, however, that if adaptive zero-knowledge is defined in this manner, then our simulation-based definition (giving M the additional power to make one oracle call to L) would imply the all-powerful party-based definition, as well.

parable to non-malleability, but which our construction also achieves, and which also suffices for constructing strong encryption schemes. The soundness property of proof systems states that with overwhelming probability, the prover should be incapable of convincing the verifier of a false statement. In this definition, we will ask that this remains the case even after a polynomially bounded party has seen a simulated proof of its choosing.

Definition 3.2 [Simulation-Sound NIZK] Let $\Pi = (f, \mathcal{P}, \mathcal{V}, \mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2))$ be an efficient non-interactive single-theorem adaptive zero-knowledge proof system for the language L. We say that Π is *simulation-sound* if for all non-uniform probabilistic polynomial-time adversaries $A = (A_1, A_2)$, we have that $\Pr\left[\,\mathsf{Expt}_{A,\Pi}(k)\,\right]$ is negligible in k, where $\mathsf{Expt}_{A,\Pi}(k)$ is the following experiment:

$$\boxed{\begin{array}{l} \mathsf{Expt}_{A,\Pi}(k): \\ \quad (\Sigma, \kappa) \leftarrow \mathcal{S}_1(1^k) \\ \quad (x, \tau) \leftarrow A_1(\Sigma) \\ \quad p \leftarrow \mathcal{S}_2(x, \Sigma, \kappa) \\ \quad (x', p') \leftarrow A_2(x, p, \Sigma, \tau) \\ \quad \texttt{return true iff} \\ \qquad (p' \neq p) \texttt{ and} \\ \qquad (x' \notin L) \texttt{ and} \\ \qquad (\mathcal{V}(x', p', \Sigma) = \texttt{true}\,) \end{array}}$$

We further define two technical properties we will desire from our NIZK proof systems. The first captures the simple requirement that simulated proofs should have sufficient internal randomness that it should be very unlikely that one can predict what the output of the simulator will be beforehand. Formally, we say an NIZK proof system Π has *unpredictable simulated proofs* if for all non-uniform polynomial-time adversaries A, we have that the following experiment has a negligible probability of success:

$$\boxed{\begin{array}{l} (\Sigma, \kappa) \leftarrow \mathcal{S}_1(1^k) \\ (x, p) \leftarrow A(\Sigma) \\ p' \leftarrow \mathcal{S}_2(x, \Sigma, \kappa) \\ \texttt{return true iff } p = p' \end{array}}$$

We also define the notion that no single proof should be convincing for more than one theorem. Formally, we say an NIZK proof system Π has *uniquely applicable proofs* if for all x, p, Σ, we have that $\mathcal{V}(x, p, \Sigma) = 1$ implies $\mathcal{V}(x', p, \Sigma) = 0$ for all $x' \neq x$. The proof systems constructed in this paper will always have unpredictable simulated proofs and uniquely applicable proofs.

The Construction. We now show, assuming that one-way functions exist, how to transform any efficient non-interactive single-theorem adaptive zero-knowledge proof system $\pi = (f_\pi, \mathcal{P}_\pi, \mathcal{V}_\pi, \mathcal{S}_\pi = (\mathcal{S}_{1\pi}, \mathcal{S}_{2\pi}))$ for a language L into an adaptively non-malleable and simulation-sound non-interactive zero-knowledge proof system $\Pi = (f_\Pi, \mathcal{P}_\Pi, \mathcal{V}_\Pi, \mathcal{S}_\Pi = (\mathcal{S}_{1\Pi}, \mathcal{S}_{2\Pi}))$ for a language L. (Π will also have unpredictable simulated proofs and uniquely applicable proofs.)

The necessary additional component will be what we call a *strong* one-time signature scheme $(Gen, Sign, Ver)$, where we strengthen the usual unforgeability requirement to require that no adversary, when given a signature of a message of its choosing, can produce a different valid signature of *any* message, including the message that was already signed. Such a signature scheme can be built from any one-way function as follows: First, choose a universal one-way hash function h mapping $\{0,1\}^*$ to $\{0,1\}^k$ (such a hash function can be based on any one-way function using the construction of [R]). Then choose $2k$ strings $x_1^0, \ldots, x_k^0, x_1^1, \ldots, x_k^1$ uniformly at random from $\{0,1\}^{3k}$, and let $y_i^b = h(x_i^b)$. The verification key will be the y_i^b's and a description of h. The signing key will be the x_i^b's. To sign a message $m \in \{0,1\}^*$, one computes $u = u_1 \ldots u_k = h(m)$, and outputs $(x_1^{u_1}, \ldots, x_k^{u_k})$. To verify a signature $(z_1, \ldots, z_k)$ on message m, one simply computes $u = h(m)$, and verifies that $h(z_i) = y_i^{u_i}$ for all i. It is straightforward to verify that this scheme has the properties we desire, and the details are skipped here. Let us assume that the public verification key VK produced by $Gen(1^k)$ is bounded in length by a polynomial $q(k)$.

We also assume there is a known efficiently computable function $g : \{0,1\}^{q(k)} \to 2^{[q'(k)]}$ mapping $q(k)$ bit strings to *distinct* subsets of $[q'(k)] = \{1, 2, \ldots, q'(k)\}$ containing precisely $q'(k)/2$ elements. For instance, one such g could be gotten by letting $q'(k) = 2q(k)$, and defining $g(x)$ to be the subset of $[q'(k)]$ that contains $2i$ if $x_i = 0$ and $2i - 1$ if $x_i = 1$.

Intuition. Dolev, Dwork, and Naor [DDN] implicitly introduced a powerful method which we call *unduplicatable set selection* using authentication mechanisms, and applied this to encryption functions. We adapt this technique to apply it to NIZK, and show that it can be used to achieve non-malleability here, as well. Furthermore, by using this in conjunction with a particular combinatorial construction which we can realize using polynomials over finite fields, we show how to achieve non-malleable NIZK for many proofs, if a polynomial bound on the number of proofs is known beforehand. We note that Di Crescenzo et al. [DIO] also implicitly apply unduplicatable set selection to attack the problem of non-malleable commitment, but do so in a complicated way. The techniques in our work can be used to provide an alternative, simpler construction to theirs.

The additional resource of the strong one-time signature scheme will be used to implement this notion of unduplicatable set selection: We choose a verification key/signing key pair (VK, SK), and then use $g(VK)$ to select a set of some objects. We then use SK to sign whatever we do with these objects, but keep SK hidden. To see why we call this unduplicatable set selection, consider what happens if some other party tries to use the *same* set of objects, but tries to do something different with them. By the properties of g, it must use the same verification key VK. By the security of the signature scheme, however, it will be unable to produce a valid signature unless it merely replicates what it already saw.

The idea will be to have many reference strings for π, and use unduplicatable set selection to select subsets of these reference strings used to prove the desired statements (*i.e.* our "objects" will be reference strings, and what we do with these objects is use them to build proofs of some fixed theorem). To simulate a proof, one needs to only select a *subset* of the reference strings to come from the simulator, while the rest can be truly random. But now, by the property of unduplicatable set selection, if the adversary is able to produce a different proof, it must have used a different set of reference strings, including at least one truly random reference string. Hence, intuitively, we can produce a proof without any help by simply using the adversary with a simulated proof, and then outputing the proof it must produce with respect to one of the truly random reference strings.

We now formalize this intuition and define Π:

- **[Reference String Length]** $f_\Pi(k) = q'(k) \cdot f_\pi(k)$. We think of the new reference string as consisting of $q'(k)$ reference strings for π, *i.e.* $\Sigma = \sigma_0 \circ \sigma_1 \circ \cdots \sigma_{q'(k)}$.
- **[Prover]** $\mathrm{P}_\Pi(x, w, \Sigma)$:
  (1) Run $Gen(1^k)$ to obtain a verification key / signing key pair (VK, SK) for the one-time signature scheme.
  (2) For each i in the set $g(VK)$, obtain $p_i = \mathrm{P}_\pi(x, w, \sigma_i)$. For $i \notin g(VK)$, let $p_i = \epsilon$, the empty string.
  (3) Let $\bar{p} = p_1 \circ p_2 \circ \cdots \circ p_{q'(k)}$.
  (4) Output $(VK, x, \bar{p}, Sign_{SK}(x, \bar{p}))$.
- **[Verifier]** $\mathcal{V}_\Pi(x, p = (VK, x', \bar{p}, z), \Sigma)$:
  (1) Check $x = x'$, and validity of one-time signature z *i.e.* $Ver_{VK}((x, \bar{p}), z) = \texttt{true}$.
  (2) Decompose $\bar{p}$ into p_i for i in $g(VK)$.
  (3) For each i in $g(VK)$, verify the proof p_i, *i.e.* $\mathcal{V}_\pi(x, p_i, \sigma_i) = \texttt{true}$.
- **[Simulator]**

$\mathcal{S}_{1\Pi}(1^k)$:
  $(VK, SK) \leftarrow Gen(1^k)$
  $(\sigma_i, \kappa_i) \leftarrow \mathcal{S}_{1\pi}(1^k)$ for $i \in g(VK)$
  $\sigma_i \leftarrow \{0,1\}^{f_\pi(k)}$ for $i \notin g(VK)$
  $\Sigma := \sigma_1 \circ \cdots \circ \sigma_{q'(k)}$
  $\texttt{return}\ (\Sigma, \kappa = (VK, SK, \{\kappa_i\}))$

$\mathcal{S}_{2\Pi}(x, \Sigma, \kappa = (VK, SK, \{\kappa_i\}))$:
  $p_i \leftarrow \mathcal{S}_{2\pi}(x, \sigma_i, \kappa_i)$ for $i \in g(VK)$.
  $p_i \leftarrow \epsilon$ for $i \notin g(VK)$.
  $\bar{p} = p_1 \circ \cdots \circ p_{q'(k)}$
  $\texttt{return}\ (VK, x, \bar{p}, Sign_{SK}(x, \bar{p}))$

That Π is an efficient non-interactive single-theorem adaptive zero-knowledge proof system for L, and that it has unpredictable simulated proofs and uniquely applicable proofs is easy to verify from the construction. We now prove that this construction also achieves adaptive non-malleability:

Proof: We follow the intuition presented above. First, we define a slightly altered version π' of the proof system π. Proofs in π' are identical to those in π, except the reference string $\sigma' = \sigma_1 \ldots \sigma_{q'(k)/2}$ for π' consists of $q'(k)/2$ different reference strings for π, and a proof is considered valid if it is valid for any of these reference strings. Clearly, the soundness error of π' can only be polynomially higher than that of π; since π has negligible soundness error, we have that π' also has negligible soundness error, and thus is a non-interactive proof system for L.

We now exhibit an adversary transformer M that transforms an adversary $A = (A_1, A_2)$ into an adversary that forges a proof for the proof system π'. On input $\rho' = \rho_1 \ldots \rho_{q'(k)/2}$ (which is a reference string for π'), and given oracle access to A_1 and A_2, M simply simulates the experiment $\mathsf{Expt}_{A,R,\Pi}$ above, except that after calling $S_{1\Pi}$ to generate $\Sigma = \sigma_1 \circ \cdots \circ \sigma_{q'(k)}$, it replaces σ_{a_i} with ρ_i, where $\{a_1, \ldots, a_{q'(k)/2}\} = \{1, \ldots, q'(k)\} \setminus g(VK)$. Since the input distribution to M is uniform, the resulting distribution on Σ is identical to the distribution output normally by $S_{1\Pi}$, and by construction $S_{2\Pi}$ will work precisely as before.

Suppose that $A_2(x, p, \Sigma, s)$ does output (x', p', aux) such that $p' \neq p$ yet $\mathcal{V}_\Pi(x', p', \Sigma) = \texttt{true}$ and $R(x', \mathrm{aux}) = \texttt{true}$. Now, since $p' = (VK', x', \bar{p}', z') \neq p = (VK, x, \bar{p}, z)$, this leaves two possibilities:

The first case is that $VK = VK'$, so p and p' differ in some other component. But the fact that p' passed the verification implies that A was able to produce a message/signature pair for VK different than the one given by M. If this case occurs with non-negligible probability, then we can use A to forge signatures and break the strong one-time signature scheme. We are assuming this is not possible, thus the case that $VK = VK'$ must occur with negligible probability.

On the other hand, if $VK \neq VK'$, then we know that the set $g(VK) \neq g(VK')$. This means that $\bar{p}'$ contains some valid proof p_i for $i \notin g(VK)$. Thus, M can simply output (x', p_i, aux), which is a valid proof for π'. This establishes the adaptive non-malleability of our NIZK proof system. ▌

Note that precisely the same proof shows that Π is simulation-sound, since the same reduction would show that if A is able to output false proofs with repsect to Π with non-negligible probability, then M will output false proofs with respect to π with non-negligible probability. But M receives as input only a truly random reference string for π. Hence, by the definition of soundness for π, it must be that M has only negligible probability of outputing a false proof.

We also note that this construction can be made more efficient by using a universal one-way hash function h that maps $\{0,1\}^{q(k)}$ to $\{0,1\}^k$, to have the sets selected according to $g(h(VK))$. The same analysis goes through with only minor modification, namely we must argue that $h(VK') = h(VK)$ occurs with negligible probability rather than $VK' = VK$, but this will follow directly from the one-wayness of the hash function h.

Generalizing to many proofs. The proof above shows that our construction achieves adaptive non-malleability when the adversary sees a single proof, but gives no guarantees for the case where more proofs are observed. Indeed, one can construct counterexamples where it fails against multiple proofs. Nevertheless, somewhat surprisingly, this level of security suffices for the application of building encryption schemes that are secure against adaptive chosen-ciphertext attack, even for multiple messages. However, we can explicitly build proof systems that remain non-malleable against multiple proofs, when a polynomial bound of the number of proofs is known in advance. Note that this extension is non-trivial; for instance, the natural idea of simply concatenating polynomially many reference strings to form a new reference string, and choosing a random one each time to prove a statement, does not work, since this would retain an inverse polynomial probability of having the same reference string used twice.

The framework we presented above, based on unduplicatable set selection, however, was designed so that we could extend it to the case of multiple proofs. Above, we simply wanted to ensure that the set of σ_i selected for each verification key (or hash of the verification key) was distinct, so that at least one σ_i would differ between the adversary's proof and the proof it received. Now, for any fixed polynomial bound $t(k)$ on the number of proofs that the adversary can ask for, we will need to ensure that any set selected by the adversary (which will be distinct from the $t(k)$ sets it has already seen with high probability by the property of unduplicatable set selection), there will be at least one σ_i that was not in any of the $t(k)$ sets that the adversary has already seen. This can be accomplished through the use of a combinatorial set system where no $t(k)$ sets cover any other set, which we can build efficiently using polynomials.

To accomplish our modification, we take the construction above and use a new function g, and modify f_Π accordingly. Recall that the input to g, which would be a verification key (or the hash of a verification key), has length $q(k)$, while the output of g is to be some subset of $[q'(k)]$. We will now suppress the dependence on k for notational convenience. Let $\ell = 2qt$, and assume ℓ is a prime power (otherwise take the next higher prime power). We construct the finite field $\mathbb{F}_\ell$ (which can be done efficiently). Let $q' = \ell^2 = O(q^2t^2)$, and associate $[q']$ with the set $\mathbb{F}_\ell \times \mathbb{F}_\ell$. The size of the sets output by g will be ℓ. Now, if g receives as input the bit string $m = m_0 m_1 \dots m_{(q-1)}$, we consider the polynomial $f_m = m_0 + m_1 x + m_2 x^2 + \cdots + m_{(q-1)} x^{(q-1)}$. The set output by $g(m)$ will be $\{(u, f_m(u)) : u \in \mathbb{F}_\ell\}$. Now, for any $m \neq m'$, since the degree of $f_m - f_{m'}$ is at most $q-1$, we know that f_m and $f_{m'}$ can agree on at most $q - 1 < \ell/2t$ values. Thus, for any set of t strings $m^1, \dots, m^t$ different from m,

$$\left| g(m) \setminus \bigcup_{i=1}^{t} g(m^i) \right| \geq \ell - t \cdot \left(\frac{\ell}{2t} \right) = \frac{\ell}{2}.$$

The simulation should pick t random verification key and signing key pairs ahead of time, and use simulated reference strings for the σ_i corresponding to these verification keys. Now, by the analysis above, after seeing t proofs, the adversary is forced to select a set of σ_i such that at least half of them are not ones that were involved in the t proofs the adversary has seen. Thus, it can be seen that the proof given for the original construction readily generalizes for this case.

4 Encryption Secure Against Adaptive Chosen-Ciphertext Attack

In this section, we present and prove the correctness of a simple construction of a public-key encryption scheme secure against adaptive chosen-ciphertext attack (CCA2) based on:

(1) Any semantically-secure public-key encryption scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$.

(2) An adaptively non-malleable (or simulation-sound) NIZK proof system $\Pi = (f_\Pi, \mathrm{P}, \mathcal{V}, \mathcal{S} =$

$(\mathcal{S}_1, \mathcal{S}_2)$) with unpredictable simulated proofs and uniquely applicable proofs for the language L of consistent pairs of encryptions, defined formally below:

$$L = \{(e_0, e_1, c_0, c_1) \ : \ \exists m, r_0, r_1 \in \{0,1\}^* \ : \\ c_0 = \mathcal{E}_{e_0}(m; r_0) \text{ and } c_1 = \mathcal{E}_{e_1}(m; r_1)\}$$

We note that L is certainly in NP, since the values of m, r_0, r_1 would witness membership in L, and certainly such values would always be of size polynomial in e_0, e_1, c_0, c_1.

Our scheme is a modification the original elegant scheme of Naor and Yung. The scheme of Naor and Yung is conceptually very simple: A message is encrypted using two independent semantically-secure encryption functions, and an NIZK proof is provided showing that both ciphertexts are encryptions of the same message. Unfortunately, the NIZK proof alone fails to provide security against *adaptive* chosen-ciphertext attack. We show that by simply replacing the NIZK proof with an adaptively *non-malleable* NIZK proof, one achieves full security against adaptive chosen-ciphertext attack. More precisely, the construction is as follows:

Let $\ell(k)$ be a polynomial bound on the length of messages to be encrypted. Let $t(k)$ be the induced polynomial bound on the amount of randomness needed by $\mathcal{E}$ to encrypt messages of length up to $\ell(k)$. Finally, let $q(k)$ be then the induced polynomial length of the reference string required by the proof system Π.

- $\mathcal{G}'(1^k)$: Call $\mathcal{G}(1^k)$ to generate two pairs (e_0, d_0) and (e_1, d_1) of encryption and decryption keys. Select a random reference string $\Sigma \in \{0,1\}^{q(k)}$ for Π.
    - The public key is $pk = (e_0, e_1, \Sigma)$.
    - The private key is $sk = (d_0, d_1)$.
- $\mathcal{E}'_{pk}(m)$: Choose $r_0, r_1 \leftarrow \{0,1\}^{t(k)}$. Let $c_0 := \mathcal{E}_{e_0}(m; r_0)$ and $c_1 := \mathcal{E}_{e_1}(m; r_1)$ and use $\mathcal{P}$ to generate a proof p relative to Σ that $(e_0, e_1, c_0, c_1) \in L$, using m, r_0, r_1 as the witness. Output (c_0, c_1, p).
- $\mathcal{D}'_{sk}(c_0, c_1, p)$: Use $\mathcal{V}$ to verify the correctness of p. If p is valid, output either of $\mathcal{D}_{d_0}(c_0)$ or $\mathcal{D}_{d_1}(c_1)$, chosen arbitrarily.

We now prove our main Theorem:

Theorem 4.1 *The encryption scheme $(\mathcal{G}', \mathcal{E}', \mathcal{D}')$ above is secure against* CCA2.

Proof: Our proof has the same overall structure as the proof of security found in [NY], but differs in most technical aspects. The main idea will be to transform an adaptive chosen-ciphertext attack against the new encryption scheme into a chosen-plaintext attack against the component encryption scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$. Hence we will conclude that since $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ is secure against chosen-plaintext attack, the new scheme $(\mathcal{G}', \mathcal{E}', \mathcal{D}')$ is secure against adaptive chosen-ciphertext attack.

Suppose that there were a probabilistic polynomial-time attacker $A = (A_1, A_2)$ which achieved inverse polynomial advantage $\epsilon(k)$ in a CCA2-attack against $(\mathcal{G}', \mathcal{E}', \mathcal{D}')$. From now on, to reduce the cumbersome nature of our notation, we will suppress dependence on k, but it should be clear where this dependence arises. We define two experiments that unfurl the definition of a CCA2-attack: In $\mathsf{Expt}_A(b)$, where $b \in \{0,1\}$, the attack is carried out and the challenge given to the adversary A is m_b (where m_0 and m_1 were the two messages specified by A after the first phase of the attack). Thus, by the definition of advantage in a CCA2 attack, we have that $\Pr[\mathsf{Expt}_A(1) = 1] - \Pr[\mathsf{Expt}_A(0) = 1] \geq \epsilon$.

We also define $\mathsf{Expt}_A^S(b_0, b_1)$, where $b_0, b_1 \in \{0,1\}$, in which the attack is carried out by a simulator – now the challenge is a ciphertext that consists of encryptions of m_{b_0} and m_{b_1}, and a simulated proof of consistency. Note that b_0 need not equal b_1, since the simulator does not need a witness to produce a proof. Formally, $\mathsf{Expt}_A^S(b_0, b_1)$ is as follows:

$\mathsf{Expt}_A^S(b_0, b_1)$:
  Set up pk, sk:
$\star$ $(\Sigma, \kappa) \leftarrow \mathcal{S}_1(1^k)$
    $(e_0, d_0) \leftarrow \mathcal{G}(1^k)$; $(e_1, d_1) \leftarrow \mathcal{G}(1^k)$
    $pk := (e_0, e_1, \Sigma)$; $sk := (d_0, d_1)$
  $(m_0, m_1, \tau) \leftarrow A_1^{\mathcal{D}'_{sk}}(pk)$
  Set up challenge:
    $r_0, r_1 \leftarrow \{0,1\}^{t(k)}$
    $c_0 := \mathcal{E}_{e_0}(m_{b_0}; r_0)$; $c_1 := \mathcal{E}_{e_1}(m_{b_1}; r_1)$
$\star$ $p \leftarrow \mathcal{S}_2((e_0, e_1, c_0, c_1), \Sigma, \kappa)$
    $y := (c_0, c_1, p)$
  $g \leftarrow A_2^{\mathcal{D}'^{(y)}_{sk}}(y, \tau)$
  `return` g

Note that the only lines that differ between $\mathsf{Expt}_A^S(b, b)$ and $\mathsf{Expt}_A(b)$ are the ones marked with a $\star$ above. For $\mathsf{Expt}_A(b)$, these would be replaced by $\Sigma \leftarrow \{0,1\}^{q(k)}$ and $p \leftarrow \mathcal{P}((e_0, e_1, c_0, c_1), (m_b, r_0, r_1), \Sigma)$, respectively.

Since $\Pr[\mathsf{Expt}_A(1) = 1] - \Pr[\mathsf{Expt}_A(0) = 1] \geq \epsilon$, it must be the case that one of the following four quantities is at least $\epsilon/4$:

$$\left| \Pr[\mathsf{Expt}_A(1) = 1] - \Pr\left[\mathsf{Expt}_A^S(1,1) = 1\right] \right| \quad (1)$$

$$\left|\Pr\left[\mathsf{Expt}_A^S(1,1)=1\right]-\Pr\left[\mathsf{Expt}_A^S(0,1)=1\right]\right| \quad (2)$$

$$\left|\Pr\left[\mathsf{Expt}_A^S(0,1)=1\right]-\Pr\left[\mathsf{Expt}_A^S(0,0)=1\right]\right| \quad (3)$$

$$\left|\Pr\left[\mathsf{Expt}_A^S(0,0)=1\right]-\Pr\left[\mathsf{Expt}_A(0)=1\right]\right| \quad (4)$$

It is easily seen that if either (1) or (4) were at least $\epsilon/4$, then this would imply a distinguisher for the simulator for Π. This leaves only the two cases of either (2) or (3) being at least $\epsilon/4$. To analyze these cases, we first define some important concepts and prove a critical lemma. We define a ciphertext $c = (c_0, c_1, p)$ to be *valid* with respect to a public key $pk = (e_0, e_1, \Sigma)$ if $\mathcal{V}((e_0, e_1, c_0, c_1), p, \Sigma) = \texttt{true}$. Note that only valid ciphertexts are ever decrypted. We define a ciphertext c to be *proper* with respect to a public key pk if $(e_0, e_1, c_0, c_1) \in L$, *i.e.* the ciphertexts c_0 and c_1 are encryptions of the same message.

The central observation now is that if the adversary makes no *improper but valid* queries to the decryption oracle during the attack, then the decryption mechanism needs *only one* of the decryption keys d_0 or d_1 in order to answer all queries made by the adversary, since all the valid queries are two encryptions of the same message. In this case, we will show how to mount a chosen-plaintext attack on the underlying encryption scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$ by simulating a chosen-ciphertext attack with the adversary and using it to break the underlying encryption scheme. We will fill in the details shortly.

For these ideas to work, however, we need to ensure that the adversary cannot make improper but valid queries. The relevant experiments here are $\mathsf{Expt}_A^S(1,1)$, $\mathsf{Expt}_A^S(0,1)$, and $\mathsf{Expt}_A^S(0,0)$. Note that in the case of $\mathsf{Expt}_A^S(0,1)$, the adversary is given an improper but valid challenge ciphertext, and yet we seek to ensure that it will not be able to produce any other such ciphertexts. Here we see that the non-malleability of the NIZK proof system Π will be critical in denying the adversary the ability to produce valid improper ciphertexts, *even* after it has seen such a ciphertext. We establish the following lemma:

Lemma 4.2 *For all $b_0, b_1 \in \{0,1\}$, and all non-uniform polynomial-time adversaries $A = (A_1, A_2)$, the probability over the experiment $\mathsf{Expt}_A^S(b_0, b_1)$ that A will make, in either stage A_1 or A_2, a valid but improper query to the decryption oracle (different from the challenge ciphertext y) is negligible in k.*

Proof: This lemma follows from the simulation soundness (or similarly from adaptive non-malleability) of Π. We build the following machines, which will be plugged into the definition of simulation soundness:

```
A'_1(Σ):
  Initialize c' := ⊥
  Set up pk, sk:
    (e_0, d_0) ← G(1^k) ; (e_1, d_1) ← G(1^k)
    pk := (e_0, e_1, Σ) ; sk := (d_0, d_1)
  Simulate first stage of attack:
    (m_0, m_1, τ_1) ← A_1^{D'_sk}(pk) where any queried
    valid improper ciphertext is stored in c'
  Set up challenge encryptions:
    r_0, r_1 ← {0,1}^{t(k)}
    c_0 := E_{e_0}(m_{b_0}; r_0) ; c_1 := E_{e_1}(m_{b_1}; r_1)
  return (x = (e_0, e_1, c_0, c_1), τ = (τ_1, d_0, d_1, c'))
```

Above, A_1' implements $\mathcal{D}'_{sk}$ for A_1, and at the same time whenever A_1 presents a query ciphertext $y' = (c_0', c_1', p')$, if y' is valid, A_1 also checks whether y' is proper by checking that $\mathcal{D}_{d_0}(c_0') = \mathcal{D}_{d_1}(c_1')$. If this is not the case, we let $c' = y'$.

The simulator will provide the proof of consistency for the two challenge encryptions computed by A_1'. We then build A_2' to complete the simulation of the second stage of the attack, again looking out for valid improper queries:

```
A'_2(x = (e_0, e_1, c_0, c_1), p, Σ, τ = (τ_1, d_0, d_1, c')):
  y := (c_0, c_1, p)
  Simulate second stage of attack:
    g ← A_2^{D'^{(y)}_sk}(y, s_1) where any queried
    valid improper ciphertext is stored in c'
  if c' = ⊥ then abort
    else return (x' = (e_0, e_1, c'_0, c'_1), p')
```

Above, A_2' implements $\mathcal{D}'^{(y)}_{sk}$ for A_2, and again simultaneously checks as above to see if A_2' makes a valid improper query, and if so lets c' be that query $y' = (c_0', c_1', p')$.

We plug the above two machines A_1' and A_2' into the definition of simulation soundness. Now, first we note that if A_1' finds a valid improper query $c' = (c_0', c_1', p')$, *i.e.* a valid improper ciphertext is found *before* the challenge ciphertext $y = (c_0, c_1, p)$ is given, then by unpredictability of simulated proofs, the probability that the proof p' found by the adversary is identical to the the proof p output by the simulator is negligible (because of the independent randomization used by the simulator). On the other hand, if A_2' finds a valid improper ciphertext $c' \neq y$, since Π has uniquely applicable proofs yet p' passes the validity test, the proof components of c' and y must differ (because we assume no proof can be convincing for two different theorems). Thus we see that the probability that $p' \neq p$, $x' \notin L$, and $\mathcal{V}(x', p', \Sigma) = \texttt{true}$ is at least the probability that A makes a valid improper query less something negligible.

But the definition of simulation soundness implies that the former probability is negligible, and hence the latter must be negligible as well.

Note that since given the decryption keys one can efficiently check the properness of a ciphertext, this argument applies with the definition of adaptive non-malleability as well: the only changes required are that A_2' should output aux $= (d_0, d_1)$ and the relation $R((e_0, e_1, c_0, c_1), (d_0, d_1))$ should be true iff c_0 and c_1 decrypt to different messages. ▌ (End of Proof of Lemma 4.2)

Now we are ready to show how to mount a chosen-plaintext attack on the semantically-secure encryption scheme $(\mathcal{G}, \mathcal{E}, \mathcal{D})$. Let us consider the case that $\Pr\left[\mathsf{Expt}_A^S(1,1) = 1\right] - \Pr\left[\mathsf{Expt}_A^S(0,1) = 1\right] \geq \epsilon/4$. (The other case folows by an exactly parallel argument.)

By the lemma, we may assume that the adversary will never make an improper but valid query. Hence, a single decryption key will suffice to implement the decryption oracle for queries made by the adversary. Hence, we may build the following chosen-plaintext attacker $B = (B_1, B_2)$:

$B_1(e)$:
  Set up pk, sk:
    $\Sigma \leftarrow \mathcal{S}_1(1^k)$
    $e_0 := e$; $(e_1, d_1) \leftarrow \mathcal{G}(1^k)$
    $pk := (e_0, e_1, \Sigma)$; $sk := d_1$
  $(m_0, m_1, s_1) \leftarrow A_1^{\mathcal{D}'_{sk}}(pk)$
  `return` (m_0, m_1, Δ)

$B_2(c, \Delta)$:
  Set up challenge:
    $r_1 \leftarrow \{0,1\}^{t(k)}$
    $c_0 := c$; $c_1 := \mathcal{E}_{e_1}(m_1; r_1)$
    $p \leftarrow \mathcal{S}_2((e_0, e_1, c_0, c_1), \Sigma)$
    $y := (c_0, c_1, p)$
  $g \leftarrow A_2^{\mathcal{D}'^{(y)}_{sk}}(y, s_1)$
  `return` g

Above, Δ stores all the necessary state needed to be transferred from B_1 to B_2, *i.e.* $e_0, e_1, d_1, \Sigma, s_1$, and the state information needed by the simulator. As in the proof of the lemma above, B_1 and B_2 implement the decryption oracle for A_1 and A_2, but because of the lemma, the single decryption key d_1 suffices. Thus, we have that B will achieve an advantage only negligibly smaller than $\epsilon/4$ in its plaintext attack on $(\mathcal{G}, \mathcal{E}, \mathcal{D})$, which we assumed is impossible. An exactly parallel argument holds for the case when $\Pr\left[\mathsf{Expt}_A^S(0,1) = 1\right] - \Pr\left[\mathsf{Expt}_A^S(0,0) = 1\right] \geq \epsilon/4$, except in this case B knows d_0 and the first component of the challenge to A_2 is always an encryption of m_0. Thus, the advantage $\epsilon(k)$ of any adaptive chosen-ciphertext attacker must be negligible, and the security of our encryption scheme is established. ▌

Remark 4.3 We note that a standard hybrid argument shows that any encryption scheme secure against adaptive chosen-ciphertext attack as defined here is also secure if the adversary is given many encryptions of the challenge message. While one would certainly hope that this is the case, in this case it is particularly surprising, since the non-interactive proof system used here is only adaptively non-malleable and zero-knowledge for a *single theorem.*

5 Conclusions

In this paper we motivated and introduced the notion of non-malleable NIZK, showed how to achieve it against any fixed number of proofs, and constructed a new simple encryption scheme based on general assumptions secure against adaptive chosen-ciphertext attack based on this notion. As argued in the introduction, we believe that non-malleable NIZK comes much closer to achieving our intuitive notion of "zero-knowledge" for non-interactive proof systems, and hence will find many other applications.

We finish with a couple of open problems. A major problem left open is how to achieve non-malleable NIZK proof systems that are secure against an unbounded number of proofs. Another question concerns our definition of non-malleability for NIZK (Definition 3.1), in which the second experiment allowed the adversary to give a proof using a possibly different non-interactive proof system. While this does capture the right semantics (since being able to prove a theorem without outside help should imply knowledge of a witness regardless of what proof system one uses), it may be useful to have a construction in which the adversary in the second experiment uses the same proof system. This would ensure a higher level of "knowledge-tightness" (the current definition allows for a polynomial loss), and could be needed in proofs of other constructions that utilize non-malleable NIZK.

Acknowledgments

The author wishes to thank Shafi Goldwasser for providing the impetus for this work by suggesting that there should be simpler way to achieve full chosen-ciphertext security under general assumptions than the construction given in [DDN], as well as for providing a great deal of assistance. We also thank Salil Vadhan for helpful conversations and ideas early in this research. The author gratefully thanks Cynthia Dwork for introducing him to the notion of non-malleability and convincing him of its importance. Finally, the author thanks Oded Goldreich for many helpful comments on the write-up.

References

[BCK] M. BELLARE, R. CANETTI AND H. KRAWCZYK, A modular approach to the design and analysis of authentication and key exchange protocols. *Proceedings of the* 30th *Annual Symposium on Theory of Computing*, ACM, 1998.

[BDPR] M. BELLARE, A. DESAI, D. POINTCHEVAL AND P. ROGAWAY, Relations among notions of security for public-key encryption schemes. *Advances in Cryptology – Crypto* 98 *Proceedings*, Lecture Notes in Computer Science Vol. 1462, H. Krawczyk ed., Springer-Verlag, 1998.

[BG] M. BELLARE, S.GOLDWASSER, New paradigms for digital signatures and message authentication based on non-interactive zero knowledge proofs. In G. Brassard, editor, Advances in Cryptology – CRYPTO '89, volume 435 of Lecture Notes in Computer Science, pages 194-211, 20-24 August 1989. Springer-Verlag, 1990.

[BR1] M. BELLARE AND P. ROGAWAY, Random oracles are practical: a paradigm for designing efficient protocols. *First ACM Conference on Computer and Communications Security*, ACM, 1993.

[BR] M. BELLARE AND P. ROGAWAY, Optimal asymmetric encryption – How to encrypt with RSA. *Advances in Cryptology – Eurocrypt* 94 *Proceedings*, Lecture Notes in Computer Science Vol. 950, A. De Santis ed., Springer-Verlag, 1994.

[BS] M. BELLARE AND A. SAHAI, Non-Malleable Encryption: Equivalence between Two Notions, and an Indistinguishability-Based Characterization, To appear, CRYPTO '99.

[BDMP] M. BLUM, A. DE. SANTIS, S. MICALI AND G. PERSIANO, Non-Interactive Zero-Knowledge Proofs. *SIAM Journal on Computing*, vol. 6, December 1991, pp. 1084–1118.

[BFM] M. BLUM, P. FELDMAN AND S. MICALI, Non-interactive zero-knowledge and its applications. *Proceedings of the* 20th *Annual Symposium on Theory of Computing*, ACM, 1988.

[CS] R. CRAMER AND V. SHOUP, A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. *Advances in Cryptology – Crypto* 98 *Proceedings*, Lecture Notes in Computer Science Vol. 1462, H. Krawczyk ed., Springer-Verlag, 1998.

[DIO] G. DI CRESCENZO, Y. ISHAI, AND R. OSTROVSKY, Non-Interactive and Non-Malleable Commitment. *Proceedings of the* 30th *Annual Symposium on Theory of Computing*, ACM, 1998.

[DP] A. DE SANTIS AND G. PERSIANO, Zero-knowledge proofs of knowledge without interaction. *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[DDN] D. DOLEV, C. DWORK, AND M. NAOR, Non-malleable cryptography. *Proceedings of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991. Also Manuscript, 1998. To appear, SIAM J. of Computing.

[DNS] C. DWORK, M. NAOR, AND A. SAHAI, Concurrent Zero-Knowledge. Preliminary version appeared in *Proceedings of the* 30th *Annual Symposium on Theory of Computing*, ACM, 1998. Full version in preparation.

[FLS] U. FEIGE, D. LAPIDOT, AND A. SHAMIR, Multiple non-interactive zero knowledge proofs based on a single random string (extended abstract). In *31st Annual Symposium on Foundations of Computer Science*, volume I, pages 308–317, St. Louis, Missouri, 22–24 October 1990. IEEE.

[Go2] O. GOLDREICH, Foundations of cryptography. Class notes, Spring 1989, Technion University.

[Go] O. GOLDREICH, A uniform complexity treatment of encryption and zero-knowledge. *Journal of Cryptology*, Vol. 6, 1993, pp. 21-53.

[GM] S. GOLDWASSER AND S. MICALI, Probabilistic encryption. *Journal of Computer and System Sciences*, 28:270–299, 1984.

[GMR] S. GOLDWASSER, S. MICALI, AND C. RACKOFF, The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, February 1989.

[NY] M. NAOR AND M. YUNG, Public-key cryptosystems provably secure against chosen ciphertext attacks. *Proceedings of the* 22nd *Annual Symposium on Theory of Computing*, ACM, 1990.

[RS] C. RACKOFF AND D. SIMON, Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack. *Advances in Cryptology – Crypto* 91 *Proceedings*, Lecture Notes in Computer Science Vol. 576, J. Feigenbaum ed., Springer-Verlag, 1991.

[R] JOHN ROMPEL, One-way functions are necessary and sufficient for secure signatures. *Proceedings of the* 22nd *Annual Symposium on Theory of Computing*, ACM, 1990.

[SET] SETCO (Secure Electronic Transaction LLC), The SET standard — book 3 — formal protocol definitions (version 1.0). May 31, 1997. Available from `http://www.setco.org/`

[Sho98] VICTOR SHOUP, Why chosen ciphertext security matters, IBM Research Report RZ 3076, November, 1998.