Non-malleable Zero-Knowledge Arguments with Lower Round Complexity

Zhenbin Yan* Yi Deng†

Abstract

Round complexity is one of the fundamental problems in zero-knowledge proof systems. Non-malleable zero-knowledge (NMZK) protocols are zero-knowledge protocols that provide security even when man-in-the-middle adversaries interact with a prover and a verifier simultaneously. It is known that *constant-round public-coin* NMZK Arguments for NP can be constructed by assuming the existence of *collision-resistant hash functions* (Pass and Rosen STOC'05) and the *four-round private-coin* NMZK Arguments for NP can be constructed in the plain model by assuming the existence of *one-way functions* (Goyal, Richelson, Rosen and Vald FOCS'14 and Ciampi, Ostrovsky, Siniscalchi and Visconti TCC'17).

In this paper, we present a *six-round public-coin* NMZK *argument of knowledge* system assuming the existence of collision-resistant hash functions and a *three-round private-coin* NMZK *argument system* from *multi-collision resistance of hash functions* assumption in the keyless setting.

---

[*] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, China. School of Cyber Security, University of Chinese Academy of Sciences, China. Email: `yanzhenbin@iie.ac.cn`

[†] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, China. School of Cyber Security, University of Chinese Academy of Sciences, China. Email: `deng@iie.ac.cn`

1 Introduction

The fundamental notion of zero-knowledge was introduced by Goldwasser et al. [GMR89]. We say an interactive proof protocol for **NP** is *zero-knowledge* if the prover can prove a statement $x \in L$ is true without revealing other useful information. With such an intriguing nature, zero-knowledge proof has played a central role in the design and study of cryptographic protocols.

The notion of *non-malleable* zero knowledge (NMZK) was first introduced and achieved by Dolev et al. [DDN00], considering the execution of zero-knowledge proofs in the setting where a *man-in-the-middle* adversary interacts with an honest prover in the left session and an honest verifier in the right session. Pass and Rosen [PR05b, PR05a] constructed the first constant-round *public-coin* NMZK arguments based on the existence of collision-resistant hash functions. Their protocols rely on the non-black-box techniques used by Barak [Bar01] to obtain constant-round public-coin ZK arguments for **NP**. We note that almost all known techniques for achieving *non-malleable* zero knowledge have relied crucially on *non-malleable* commitments. However, Pass and Rosen [PR05b, PR05a] reversed the roles. They first achieved a constant-round NMZK protocol and then used it to achieve non-malleable commitment schemes and *concurrent* non-malleable commitment schemes. For the case of *private-coin* zero-knowledge, Goyal et al. [GRRV14] used an algebraic approach and achieved the first 4-round NMZK argument based on the minimal assumption that *one way functions* (OWFs) exist. Very recently, Ciampi et al. [COSV17a] achieved a *delayed-input* 4-round NMZK argument based on OWFs. In this work, we continue along this line of research and mainly focus on the following two problems:

Problem 1. *We know that* public-coin *interactive proof protocol has many advantages, such as public verifiability, better resilience to leakage and so on. The round complexity of the original public-coin ZK protocols of [Bar01] is six rounds. In this paper, we focus on how to reduce the round complexity of the current known* public-coin non-malleable zero-knowledge *protocols, (i.e., [PR05b, PR05a]). In particular, we provide a construction of such protocol in six rounds based on standard collision-resistant hash functions assumptions. This demonstrates that, like* private-coin *zero-knowledge arguments,* public-coin non-malleable *zero-knowledge arguments can achieve the same round complexity as public-coin zero-knowledge arguments without the cost of additional complexity assumptions.*

Theorem 1.1. *Assuming the existence of collision-resistant hash functions, there exists a 6-round public-coin non-malleable zero-knowledge argument of knowledge system for* ***NP****.*

As a corollary of our first result, using Barak et al.'s transformation [BGGL01], we can modify our construction into a resettably-sound non-malleable zero-knowledge arguments. This can be done by having the verifier V sample a seed s for a pseudorandom function (PRF) f_s at the beginning of the protocol, and generate each of its message by applying f_s to the current history message it received. Thus, from [BGGL01] we can get the following result.

Theorem 1.2. *Assuming the existence of collision-resistant hash functions, there exists a 6-round non-malleable zero-knowledge argument system for* ***NP*** *that is also resettably-sound.*

Problem 2. *Ever since the introduction of zero-knowledge, an extensive amount of research has been dedicated towards minimizing their round complexity. In the negative direction,* two-round *ZK arguments for NP were ruled out by Goldreich and Oren [GO94] and* three-round black-box *ZK arguments for NP were ruled out by Goldreich and Krawcyzk [GK90]. In the positive direction, three-round ZK arguments with non-black-box simulators were studied in [BBK$^+$16, BCPR16]* et al.*, and* *very recently, Bitansky, Kalai, and Paneth [BKP18] constructed general* three-round *ZK arguments for* non-uniform *provers and verifiers based on* keyless multi-collision-resistant hash functions*.*

In this paper, we further study how to construct the 3-round non-malleable zero-knowledge arguments *for **NP** based on the* multi-collision resistance of hash functions*. We present the following results:*

Theorem 1.3. *Assuming keyless multi-collision resistant hash functions, LWE and DDH (or QR or N^{th} residuosity), all quasi-polynomially hard, there exists 3-round non-malleable zero-knowledge arguments for **NP**.*

1.1 Technique Overview

Public-coin NMZK argument. Recall that due to the impossibility of *constant-round public-coin zero-knowledge proofs* with black-box simulator proved by Goldreich and Krawczyk [GK90, GK96], all currently known *public-coin zero-knowledge* protocol such as [Bar01, CLP13a, CLP13b, PRT13, PTW09] et al. are constructed by using Barak's non-black-box simulation technique [Bar01]. In order to reduce the round complexity, our protocol will try to combine a variant of Barak's protocol (i.e., the "special-purpose" universal argument [PR05b]) and a 3-round *public-coin extractable non-malleable* commitment scheme [GPR16] in parallel. We let the prover commit to a witness of the statement by using a non-malleable commitment scheme and proves that it either committed to a valid witness or a valid trapdoor by using the "special-purpose" universal argument protocol.

The "special-purpose" universal argument [PR05b] consists of three stages. The first stage is a commitment-challenge slot, the second stage is an "*encrypted*" universal argument to argue that the prover has a trapdoor, and the last stage is a 3-round public-coin WIPOK to prove that either it knows a witness for statement $x \in L$ or the opening of the transcript in stage two is an accepting proof for knowing the trapdoor. We observe that if we directly combine the 3-round of the WIPOK subprotocol with a 3-round non-malleable commitment subprotocol in parallel, then there is a subtle issue that we cannot deal with. We explain informally the difficulties that we encounter in the following.

Recall that the definition of non-malleable zero-knowledge requires the existence of a simulator-extractor SE that can simulate the view of a *man-in-the-middle* adversary $\mathcal{A}$ while simultaneously extracting the witnesses of the statement proved by the adversary in the right interaction. On the one hand, in order to prove the zero-knowledge property, we need the non-malleable commitment to commit to a witness for statement $x \in L$ in real execution or a trapdoor in the simulator execution. On the other hand, in order to prove the non-malleable property, we need to argue that when the simulator commits a trapdoor (instead of witnesses) in the left interaction, the *man-in-the-middle* adversary $\mathcal{A}$ still cannot change its committed values in the right interaction. The usual approach is to design a series of hybrids and reduce the security to the *witness indistinguishability* of the WIPOK on the left. However, when we extract the witnesses on the right using the rewinding method, the WIPOK subprotocol (i.e. [Blu86]) on the left side will also be rewound, because the adversary can use a scheduling to make the two sides execute in parallel. Thus, we cannot use the simulator-extractor to break the witness indistinguishability of the WIPOK and arrive at a contradiction.

We bypass this obstacle by using the *public-coin* non-malleable commitment scheme to commit the witness *twice in sequence*. More specifically, we execute the two non-malleable commitments in five rounds where the messages in the third round of the first commitment and the first round of the second commitment are sent in one round. We let the *five rounds* non-malleable commitments execute *in parallel with* the *six rounds* "special-purpose" universal argument from its second round, and keep the round complexity at six (see Fig. 1). Using this special structure, *no matter what* scheduling strategy the adversary uses, we can design a series of hybrids to complete the security proof. Because the simulator now has the freedom to choose which non-malleable commitment to extract the witness. (For the details see Sec. 3.1).

Three-round NMZK argument. Due the lower bound of the three-round black-box zero-knowledge showed by Goldreich and Krawczyk [GK90, GK96], three-round zero-knowledge protocols necessarily require using the code of the adversary in a non-black-box way. But Barak's non-black-box technique [Bar01] does not apply to this question, because the two-round trapdoor statement generation steps must be fixed before the three-round WI universal argument and its round complexity exceeds three rounds. Another problem is that the protocol is only three rounds, we can not invoke the non-malleable commitment twice in sequence as before to bypass the reduction security to the WI subprotocol. In order to squeeze the round complexity, here we consider two main components to achieve our three rounds NMZK argument.

The first component is the variant of the *two-message memory delegation technique* [CKLR11] based on the LWE assumption. This approach first appeared in Bitansky et al. [BBK$^+$16, BKP18] to construct three-round zero-knowledge argument based on the existence of *keyless multi-collision resistant hash functions* and LWE assumptions against quasipolynomial-time attacker. Roughly speaking, the variant of the *verifiable memory delegation protocol* consists of three rounds. In the first round, the prover provides the verifier a short commitment to the digest d of a large memory D. In the second round, the verifier sends a query q to the prover and holds a secret state vst itself. The verifier can then delegate any arbitrary deterministic computation M to be executed over the memory D. In the third round, the prover responds with the Turing machine M, the computation's output y, the committed values d and t, as well as a short proof π for the correctness that $\exists$ D such that $M(D) = y$ within t steps where d is the digest of the memory D. The verifier can verify the computation by using its secret state vst in time that is independent of that of the delegated computation and the size of the memory. Inspired by Barak's non-black-box simulation idea when considering the construction of three-round zero-knowledge arguments, Bitansky et al. [BBK$^+$16, BKP18] see the trapdoor (i.e. the verifier's code V^*) as the memory. The simulator can then prove that $M(V^*) = r$ *in one round* by using the *verifiable memory delegation protocol* above. In particular, the simulator can construct a Turing machine M such that on input the memory $D = V^*$, it parses V^* as a Turing machine which on input c will output the challenge string r.

The second component is the *delayed-input statistical witness indistinguishable arguments* (SWI) in two rounds. Very recently, Kalai, Khurana and Sahai [KKS18] first achieved this protocol based on *quasi-polynomial hardness of two-message oblivious transfer*, which in turn can be based on the DDH or QR or N^{th} residuosity assumption against quasi-polynomial attacker. Such protocol has an advantage that the witness indistinguishable arguments will not leak the information about which witness is used by the simulator, because it is statistically secure. Therefore, when the simulator changes the witness used in the SWI from the real witness to the trapdoor on the left side, it will not affect the values of the adversary commitment on the right side. We note that the assumption of using statistical WI is stronger than using standard WIPOK, i.e., the Blum [Blu86] protocol or the FLS [FLS99] protocol before. However, this is not an issue, because our first component has been based on the quasi-polynomial hard assumption.

Thus, in our non-malleable zero-knowledge arguments, we can combine the *two-round* SWI [KKS18], the *three-round verifiable memory delegation protocol* [BKP18] and the *three-round public-coin extractable non-malleable* commitment scheme [GPR16] in parallel. More specifically, the prover runs the two-round SWI protocol to prove that either it knows the opening of the non-malleable commitment to w such that $R(x, w) \in R_L$ or knows a short proof π for the correctness of the delegate computation $M(V^*) = r$.

It is easy to see that the *zero-knowledge property* of the above protocol can be directly obtained from the original protocol [BKP18]. The *non-malleable property* can be obtained from the *non-malleable security* of the NMCom, the *statistical witness-indistinguishable* of SWI and the *computational hiding* of the *non-interactive primitive* Com. In particular, there are only *three rounds* in the protocol. The malicious scheduling the man-in-the-middle adversary can be used only two types, i.e., *the synchronous scheduling* where he lets the right interaction run *in parallel* with the execution of the left or *the sequential scheduling* where he lets the right interaction complete after the execution of the left. We can give a series of hybrid proofs separately to argue the *non-malleable property*. (For the details see Sec. 3.2).

1.2 Related Work

The round complexity of *non-malleable zero-knowledge* and the *non-malleable commitment* are usually studied in parallel. In particular, the former relies heavily on the latter. Pass [Pas13, Pas16] showed a lower bounds of the black-box reductions from *two-round* non-malleable commitment to any *standard intractability assumptions*. Since then, Goyal et al. [GRRV14, GPR16] constructed a *four-round* non-malleable commitment based on the existence of *one-way function* and a *three-round protocol* using *quasi-polynomially hard injective one-way functions*. Recently, Ciampi et al. [COSV16, COSV17b] showed a *transformation* from any *three-round* non-malleable commitment ([GPR16]) to *three-round* concurrent non-malleable commitment and constructed a *four-round* concurrent non-malleable commitment based on the existence of *one-way function*. Very recently, Lin et al. [LPS17] demonstrates the existence of a *two-round* concurrent non-malleable commitment assuming the existence of *non-interactive commitments, ZAPs, CRHFs and a time-lock puzzle* with sub-exponential security. Khurana et al. [KS17, Khu17] showed a *two-round* non-malleable commitment using *sub-exponentially hard one-way permutations, sub-exponential ZAPs, and sub-exponential DDH* and a *three-round* concurrent non-malleable commitment assuming the existence of *polynomial hardness of DDH assumption or Quadratic Residuosity or N^{th} Residuosity*, together with *ZAPs*.

The notion of *multi-collision resistance* was studied concurrently and independently by Berman et al. [BDRV18], Bitansky et al. [BKP18] and Komargodski et al. [KNY18]. Very recently, Bitansky and Lin [BL18] showed a *fully-concurrent one-message* non-malleable commitments against all efficient *non-uniform* adversaries assuming *multi-collision-resistant keyless hash functions* (and *injective OWFs*, NIWI and *time-lock puzzles*) with sub-exponential security.

1.3 Organizations

The rest of this paper is organized as follows. In Sect. 2, we give the basic definitions used throughout the paper, including the definition of non-malleable commitment, non-malleable zero-knowledge, keyless multi-collision resistant hash functions, memory delegation, and so on. Next, we describe and analyze the *six-round public-coin* non-malleable zero-knowledge argument of knowledge in Sect. 3.1 and the *three-round* non-malleable zero-knowledge in Sect. 3.2.

2 Preliminary

2.1 Notations

Let $\mathbb{N}$ denote the set of all positive integers, for any integer $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \cdots, n\}$, and let $\{0,1\}^n$ denote the set of n-bit long strings; furthermore, let μ denote a *negligible* function, if for every positive polynomial p and all *sufficiently large* n, it holds that $\mu(n) < 1/p(n)$. We assume familiarity with interactive Turing machines and interactive protocols. Let PPT denotes *probabilistic polynomial time* Turing machines, and denote by $(A(y), B(z))(x)$ an interactive protocol on the *common input* x, where A with a *private input* y, B with a *private input* z, and the random tape of each machine is *uniformly* and *independently* chosen.

2.2 Statistically Binding Commitment Schemes

The hiding property of a commitment scheme is that the sender commits to a value while keeping it *secret* from the receiver; the binding property is that the commitment can only be opened to a *single* value as determined during the commitment protocol. In the *statistically (perfectly)* binding commitments, the binding property holds against *unbounded* adversaries, while the hiding property only holds against *computationally bounded* (*non-uniform*) adversaries.

The *non-interactive* perfectly-binding commitment schemes can be constructed using any one-to-one one-way function (see Section 4.4.1 of [Gol01]). The *two-message* statistically binding commitment schemes can be obtained from any one-way function [Nao91, HILL99].

2.3 Witness Indistinguishability

Definition 1. *(Witness Indistinguishability) An interactive protocol (P, V) for $L \in \boldsymbol{NP}$ is witness indistinguishable for R_L if for every PPT adversarial verifier V^* and for every two sequences $\{w_x^1\}_{x\in L}$ and $\{w_x^2\}_{x\in L}$ such that $(w_x^1, w_x^1) \in R_L(x)$, the following ensembles are computationally indistinguishable:*

- $\{\mathit{View}_{V^*} \left\langle P(x, w_x^1) \leftrightarrow V^*(x)\right\rangle\}_{x\in L}$
- $\{\mathit{View}_{V^*} \left\langle P(x, w_x^2) \leftrightarrow V^*(x)\right\rangle\}_{x\in L}$

Definition 2. *(Proof of Knowledge [FS90, BG92]) An interactive proof (or argument) system (P, V) for an $\boldsymbol{NP}$ language L with witness relation R_L is said to be proof (or argument) of knowledge, if there exists a polynomial q, a negligible function μ, and a probabilistic oracle machine E (also called the knowledge extractor), such that for every interactive machine P^* (or PPT P^*) and every $x \in L$, the following holds: if $\Pr[(P^*, V)(x) = 1] > \mu(|x|)$, then $E^{P^*(x)}(x)$ can output a witness for x with oracle access to $P^*(x)$, and the running time of E is bounded by $\frac{q(x)}{\Pr[(P^*,V)(x)=1]-\mu(|x|)}$.*

Definition 3. *(Special soundness) A three-round interactive protocol (P, V) for $L \in \boldsymbol{NP}$ with witness relation R_L is special-sound, if there exist two accepting transcripts (α, β, γ) and $(\alpha, \beta', \gamma')$ such that the first message are the same but the challenges (β, β') are different, then there is a deterministic polynomial time algorithm which can extract the witness from the two transcripts.*

Three-round special-soundness public-coin WIPOK protocol [Blu86, FLS99] can be constructed from any one-way function. Two-round statistically witness indistinguishable arguments [KKS18] can be obtained assuming the existence of a *quasi-poly secure* OT, which can in turn be instantiated based on *quasi-poly hardness* of the *DDH* assumption [NP01], or based on the *quasi-poly hardness* of *QR* or the N'th residuosity assumption [HK12].

2.4 Non-malleable Commitment [LP11]

A *tag-based* commitment scheme $\langle C, R\rangle$ is a commitment scheme where the committer and the receiver receive a *tag* $\in \{0, 1\}^n$ (also called id) as *common input*. Consider a *man-in-the-middle* adversary $\mathcal{A}$ that, on auxiliary input z, participates in a left and a right interaction. On the left, $\mathcal{A}$ interacts with C, receiving a commitment to the value v, using identity id of its choice. On the right, $\mathcal{A}$ interacts with R attempting to commit to a related value $\widetilde{v}$, using identity $\widetilde{\mathsf{id}}$ of its choice. If $\widetilde{\mathsf{id}} = \mathsf{id}$ or the right commitment is invalid, or undefined, its value $\widetilde{v}$ is set to $\perp$. Let $\mathsf{nmc}^{\mathcal{A}}(v, z)$ denote a random variable that describes the value $\widetilde{v}$ and the view of $\mathcal{A}$ in the above experiment.

Definition 4. *(Non-Malleable Commitment) A commitment scheme $\langle C, R\rangle$ is said to be non-malleable if for every polynomial $p(\cdot)$, and every PPT man-in-the-middle adversary $\mathcal{A}$, the following ensembles are computationally indistinguishable.*

- $\{\mathsf{nmc}^{\mathcal{A}}(v, z)\}_{v,v'\in\{0,1\}^n, z\in\{0,1\}^*}$
- $\{\mathsf{nmc}^{\mathcal{A}}(v', z)\}_{v,v'\in\{0,1\}^n, z\in\{0,1\}^*}$

The *three-round public-coin synchronous honest-extractable non-malleable commitment scheme* [GPR16] can be obtained from 1-1 *one-way functions* (or, *four-round* protocol can be obtained from *one-way functions*), which is extractable w.r.t. honest sender and non-malleable against *synchronous* adversaries.

2.5 Non-malleable Zero-knowledge [LPTV10]

Let (P, V) be an interactive protocol for a language L. Consider a PPT *man-in-the-middle* adversary $\mathcal{A}$ that, given the *common input* x and an *auxiliary input* $z \in \{0, 1\}^*$. On the left, $\mathcal{A}$ acts as a verifier V^* to interact with P using the common input x and id, and the prover P will be given a valid witness $w \in R_L(x)$. On the right, $\mathcal{A}$ acts as a prover P^* that, on common input $\widetilde{x}$ to prove the validity using $\widetilde{\mathsf{id}}$. During the experiment, the statement $\widetilde{x}$ and the tags $\mathsf{id}, \widetilde{\mathsf{id}}$ are all chosen by the adversary $\mathcal{A}$. Let $\mathsf{view}_{\mathcal{A}}(1^n, x, z)$ denotes the random variable that describes the view of $\mathcal{A}$ in the above experiment. Loosely speaking, an interactive proof is a *non-malleable zero-knowledge* protocol, if for all *man-in-the-middle* adversary $\mathcal{A}$, there exists a PPT machine (called the simulator-extractor) that can simulate both the left and the right interaction for $\mathcal{A}$, while outputting a witness for the statement proved by the adversary in the right interaction.

Definition 5. *An interactive protocol (P, V) for $L \in \boldsymbol{NP}$ is said to be non-malleable zero-knowledge if for every $n \in \mathbb{N}$, and every PPT man-in-the-middle adversary $\mathcal{A}$, there exists a PPT machine SE such that:*

1. *The following ensembles are computationally indistinguishable:*

   - $\{\mathsf{view}_{\mathcal{A}}(1^n, x, z)\}_{n\in\mathbb{N}, x\in L\cap\{0,1\}^n, z\in\{0,1\}^n}$
   - $\{\mathsf{S}(1^n, x, z)\}_{n\in\mathbb{N}, x\in L\cap\{0,1\}^n, z\in\{0,1\}^n}$

   where $\mathsf{S}(1^n, x, z)$ is the first output of $\mathsf{SE}(1^n, x, z)$.

2. *Let $\widetilde{x}$ be the statements to be proved in the right interaction and $(\mathsf{view}, \widetilde{w})$ denote the output of $\mathsf{SE}(1^n, x, z)$. Then if the right interaction is accepting and $\widetilde{\mathsf{id}} \neq \mathsf{id}$, it holds that $\widetilde{w}$ is a valid witness such that $R_L(\widetilde{x}, \widetilde{w}) = 1$.*

2.6 Weak (K, γ)-Collision Resistance [BKP18]

Definition 6. *Let* $K(\cdot,\cdot)$ *be a function and* $\lambda \in \mathbb{N}$ *be a security parameter. We say that* H *is weakly* (K,γ)*-collision-resistant if for any probabilistic* $\gamma^{O(1)}$*-time* $\mathcal{A}$*(possibly* $\gamma = \lambda^{\omega(1)}$*) and any sequence of polynomial-size advice* $\{z_\lambda\}_{\lambda\in\mathbb{N}}$*, there is a negligible function* μ*, such that for any* $\lambda \in \mathbb{N}$*, letting* $K = K(\lambda, |z_\lambda|)$*,*

$$\Pr\left[\begin{array}{c} Y_1 = \cdots = Y_K \\ \forall i \neq j : X_i \neq X_j \end{array} \;\middle|\; \begin{array}{l} \mathsf{hk} \leftarrow \mathsf{H.Gen}(1^\lambda) \\ (X_1, \cdots, X_K) \leftarrow \mathcal{A}(\mathsf{hk}; z_\lambda) \\ \forall : Y_i = \mathsf{H.hash}(hk, X_i) \end{array}\right] \leq \mu(n).$$

2.7 Weak Memory Delegation [BKP18]

Definition 7. *A two-message delegation scheme* $\mathsf{MD} = (\mathsf{MD.Gen}, \mathsf{MD.Mem}, \mathsf{MD.Query}, \mathsf{MD.Prove}, \mathsf{MD.Ver})$ *satisfies:*

Correctness: *There exists a universal polynomial* $p(\cdot)$ *such that for every security parameter* $\lambda \in \mathbb{N}$*, every* $(M, t, y) \in \{0,1\}^\lambda$*, and every* D *such that* $M(D)$ *outputs* y *within* t *steps, and* $|D| \leq t \leq 2^\lambda$ *:*

$$\Pr\left[\mathsf{MD.Ver}(\mathsf{pp}, \mathsf{d}, (M, \mathsf{t}, y), \mathsf{vst}, \pi) = 1 \;\middle|\; \begin{array}{l} \mathsf{pp} \leftarrow \mathsf{MD.Gen}(1^\lambda) \\ \mathsf{d} \leftarrow \mathsf{MD.Mem}(\mathsf{pp}, D) \\ (\mathsf{q}, \mathsf{vst}) \leftarrow \mathsf{MD.Query}(1^\lambda) \\ \pi \leftarrow \mathsf{MD.Prove}(\mathsf{pp}, D, (M, t, y), \mathsf{q}) \end{array}\right] = 1,$$

where the prover $\mathsf{MD.Prove}(\mathsf{pp}, D, (M, t, y), \mathsf{q})$ *runs in time* $p(\lambda, t)$ *and the verifier* $\mathsf{MD.Ver}(\mathsf{pp}, \mathsf{d}, (M, t, y), \mathsf{vst}, \pi)$ *runs in time* $p(\lambda)$*.*

Weak Soundness for Computation-Time Bound $\bar{t}(\lambda)$**:** *For every pair of PPT adversaries* $(\mathcal{A}_1, \mathcal{A}_2)$ *and polynomial-size advice* $\{z_\lambda\}_{\lambda\in\mathbb{N}}$*, there is a negligible function* μ*, such that for every* $t(\lambda) \leq \bar{t}^{O(1)}$*, any ensemble of samplable entropic distributions* $\{Y_\lambda\}_{\lambda\in\mathbb{N}}$ *such that the min-entropy of* Y_λ *is* $\Omega(\lambda)$*, letting* $K = K(\lambda, |z_\lambda|, t)$*,*

$$\Pr\left[\mathsf{MD.Ver}(\mathsf{pp}, \mathsf{d}, (M, \mathsf{t}, y), \mathsf{vst}, \pi) = 1 \;\middle|\; \begin{array}{l} \mathsf{pp} \leftarrow \mathsf{MD.Gen}(1^\lambda) \\ (\mathsf{d}, M, \mathsf{st}) \leftarrow \mathcal{A}_1(\mathsf{pp}; z_\lambda) \\ (\mathsf{q}, \mathsf{vst}) \leftarrow \mathsf{MD.Query}(1^\lambda) \\ y \leftarrow Y_\lambda \\ \pi \leftarrow \mathcal{A}_2(\mathsf{q}, y; \mathsf{st}) \end{array}\right] \leq \mu(\lambda).$$

2.8 1-Hop Homomorphic Encryption [GHV10]

Definition 8. *A scheme* $(\mathsf{Enc}, \mathsf{Eval}, \mathsf{Dec})$*, where* $\mathsf{Enc}, \mathsf{Eval}$ *are probabilistic and* Dec *is deterministic, is a semantically-secure, circuit-private, 1-hop homomorphic encryption scheme if it satisfies the following properties:*

Perfect correctness: *For any* $n \in \mathbb{N}$*,* $x \in \{0,1\}^n$ *and circuit* C*:*

$$\Pr[(\mathsf{ct}, \mathsf{sk}) \leftarrow \mathsf{Enc}(x) : \hat{\mathsf{ct}} \leftarrow \mathsf{Eval}(\mathsf{ct}, C) \wedge \mathsf{Dec}_{\mathsf{sk}}(\hat{\mathsf{ct}}) = C(x)] = 1.$$

Semantic security: *For any non-uniform PPT* $\mathcal{A} \in \{\mathcal{A}_n\}_{n\in\mathbb{N}}$*, and any pair of inputs* $x_0, x_1 \in \{0,1\}^{poly(n)}$ *of equal length:*

$$\Pr[\mathsf{b} \leftarrow \{0,1\}, \mathsf{ct} \leftarrow \mathsf{Enc}(x_{\mathsf{b}}) : \mathcal{A}_n(\mathsf{ct}) = \mathsf{b}] \leq \frac{1}{2} + negl(n).$$

Circuit privacy: *Let $\mathcal{E}(x) = \mathsf{Supp}(\mathsf{Enc}(x))$ be the set of all legal encryptions of x, $\mathcal{E}_n = \cup_{x\in\{0,1\}^n}\mathcal{E}(x)$ be the set legal encryptions for strings of length n, and $\mathcal{C}_n$ be the set of all circuits on n input bits. There exists a (possibly unbounded) simulator $\mathsf{S}_{\mathsf{1hop}}$ such that:*

$$\{C, \mathsf{Eval}(c, C)\}_{\{n\in\mathbb{N}, \mathcal{C}\in\mathcal{C}_n, x\in\{0,1\}^n, c\in\mathcal{E}(x)\}} \overset{c}{\approx} \{C, \mathsf{S}_{\mathsf{1hop}}(c, C(x), |C|)\}_{\{n\in\mathbb{N}, \mathcal{C}\in\mathcal{C}_n, x\in\{0,1\}^n, c\in\mathcal{E}(x)\}},$$

$$\{C, \mathsf{Eval}(c, C)\}_{\{n\in\mathbb{N}, \mathcal{C}\in\mathcal{C}_n, c\notin\mathcal{E}_n\}} \overset{c}{\approx} \{C, \mathsf{S}_{\mathsf{1hop}}(c, \perp, |C|)\}_{\{n\in\mathbb{N}, \mathcal{C}\in\mathcal{C}_n, c\notin\mathcal{E}_n\}}.$$

Theorem 2.1. [BKP18] *For any (arbitrary small) $\tau(\lambda) = \omega(log\lambda)$, there exists $\bar{t}(\lambda) = \lambda^{\omega(1)}$ such that assuming a weakly (K, γ)-collision-resistant hash, for $K(\lambda, |z_\lambda|) = poly(\lambda, |z_\lambda|)$ and $\gamma(\lambda) = \lambda^\tau$, and quasi-poly$(\lambda)$-secure fully-homomorphic encryption, there exists a two-message memory-delegation scheme with weak soundness for computation-time bound $\bar{t}$.*

3 The Protocol

3.1 6-round Public-Coin Non-malleable Zero-Knowledge

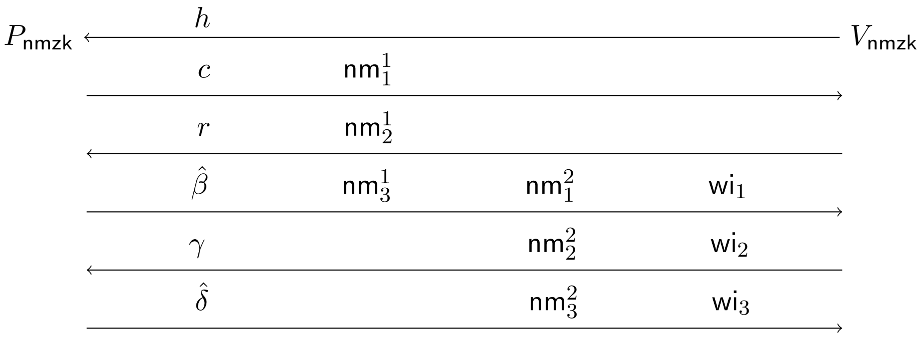
In this section we give our construction of the *6-round public-coin non-malleable zero-knowledge* protocol. We use the following building blocks:

- Non-interactive *statistically binding commitment scheme*: Com.
- 3-round *public-coin honest-extractable non-malleable commitment scheme*: NMCom.
- 3-round *"delay-input" special-soundness public-coin WIPOK* be instantiated with the FLS protocol [FLS99] : sWI.
- 6-round *"special purpose" universal argument*: sUA.

Consider a language $L \in \mathbf{NP}$ and a security parameter n, and let the prover and verifier receive a *common input* $x \in \{0,1\}^n$, $\mathsf{id} \in \{0,1\}^n$. The auxiliary input to the prover is a $\mathbf{NP}$ witness w such that $R_L(x, w) = 1$. Let $\{\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3\}$ be the transcript of the sWI, $(\mathsf{nm}_1, \mathsf{nm}_2, \mathsf{nm}_3)$ be the transcript of the commitment to the witness w computed using NMCom under tag id, and $(\hat{\beta}, \gamma, \hat{\delta})$ be the transcript of the "encrypted" UA of the sUA.

Now, we start by describing a variant of Barak's relation, which we denote by $\mathbf{R}_{\mathsf{sim}}$. Let $\mathcal{H}_n$ be a family of hash functions and $h \in \mathcal{H}_n$: $\{0,1\}^* \to \{0,1\}^n$. The relation $\mathbf{R}_{\mathsf{sim}}$ and the language L_{ua} are described in Fig. 2. Roughly speaking, we say $((h, c, r), (M, \rho_1, y)) \in R_{\mathsf{sim}}$ iff $M \in \{0,1\}^{n^{\omega(1)}}, \rho_1 \in \{0,1\}^{poly(n)}$, $|y| \leq |r| - n$ and $c = \mathsf{Com}(h(M), \rho_1)$ such that $M(y) = r$ within $n^{\omega(1)}$ steps. We say the "encrypted" UA transcript $(h, c, r, \hat{\beta}, \gamma, \hat{\delta}) \in L_{\mathsf{ua}}$, iff there exist $(M, \rho_1, y, \beta, \rho_2, \delta, \rho_3)$ such that $c = \mathsf{Com}(h(M), \rho_1)$, $\hat{\beta} = \mathsf{Com}(\beta, \rho_2)$, $\hat{\delta} = \mathsf{Com}(\delta, \rho_3)$ and $(h, \beta, \gamma, \delta)$ is an accepting transcript of universal argument proving the statement: $((h, c, r), (M, \rho_1, y)) \in R_{\mathsf{sim}}$.

Next, we give our *6-round NMZK argument* protocol description Fig. 3, and an high-level description of our *6-round public-coin NMZK argument* is described in Fig. 1. In stage one, the prover and the verifier generate the trapdoor statement (h, c, r) by running Barak's protocol and compute the first non-malleable commitment to the witness w under the identity id. More specifically, in the first round, the verifier sends a random hash function $h \overset{R}{\leftarrow} \mathcal{H}$, where $h : \{0,1\}^* \to \{0,1\}^n$. In the second round, the honest prover computes $c = \mathsf{Com}(0^n, \rho_1)$ using Com and the first round message $\mathsf{nm}_1^1(w, s_1)$ using NMCom and sends c, nm_1 to V. In the third round, the verifier computes $r \overset{R}{\leftarrow} \{0,1\}^{2n}$ and the second round message $\mathsf{nm}_2^1 \overset{R}{\leftarrow} \{0,1\}^n$ using NMCom, and sends r, nm_2^1 to P.



- $(h, c, r, \hat{\beta}, \gamma, \hat{\delta})$ is the transcript computed using the "special purpose" universal argument.
- $(\mathsf{nm}_1^1, \mathsf{nm}_2^1, \mathsf{nm}_3^1)$ is the transcript of the first commitment to the witness w computed using the honest-extractable non-malleable commitment scheme NMCom under the identity id.
- $(\mathsf{nm}_1^2, \mathsf{nm}_2^2, \mathsf{nm}_3^2)$ is the transcript of the second commitment to the witness w computed using the honest-extractable non-malleable commitment scheme NMCom under the identity id.
- $(\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3)$ is the transcript of the sWI proving the knowledge of the witness w for $x \in L$ and the opening of the two non-malleable commitments to w **or** the knowledge of the opening of the "encrypted" UA , *i.e.*, $(\hat{\beta}, \gamma, \hat{\delta})$, which is an accepting UA proof for the statement $(h, c, r) \in L_{\mathsf{sim}}$.

Figure 1: 6-round public-coin NMZK

In stage two, P and V run the three round sWI, NMCom and "encrypted" sUA in parallel, where P will compute a second non-malleable commitment to the witness w under the identity id. More specifically, in the fourth round, the honest prover computes the third round message $\mathsf{nm}_3^1(w, s_2)$ to complete the first non-malleable commitment, the first round message $\mathsf{nm}_1^2(w, s_3)$ of the second non-malleable commitment, the first "encrypted" UA message $\hat{\beta} = \mathsf{Com}(0^n, \rho_2)$ and the first sWI message wi_1 and sends $(\hat{\beta}, \mathsf{wi}_1, \mathsf{nm}_3^1, \mathsf{nm}_1^2)$ to V. In the fifth round, the verifier computes three public-coins $\mathsf{nm}_1^2, \gamma, \mathsf{wi}_2 \xleftarrow{R} \{0,1\}^n$ and sends $(\gamma, \mathsf{wi}_2, \mathsf{nm}_1^2)$ to P. In the sixth round, the honest prover computes the third round message $\mathsf{nm}_3^2(w, s_4)$ to complete the second non-malleable commitment, $\hat{\delta} = \mathsf{Com}(0^n, \rho_3)$ to complete the "encrypted" UA, wi_3 to complete the sWI and sends $(\hat{\delta}, \mathsf{wi}_3, \mathsf{nm}_3^2)$ to V.

Finally, the verifier accepts if the transcript $(\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3)$ is an accepting proof for the OR statements: $\exists(w, \mathsf{dec}_1, \mathsf{dec}_2)$ s.t. $R_L(x, w) = 1$ and (w, dec_1) is the decommitment for the transcript $(\mathsf{nm}_1^1, \mathsf{nm}_2^1, \mathsf{nm}_3^1)$ and (w, dec_2) is the decommitment for $(\mathsf{nm}_1^2, \mathsf{nm}_2^2, \mathsf{nm}_3^2)$ **or** $\exists(M, \rho_1, y, \beta, \rho_2, \delta, \rho_3)$ s.t. $(h, c, r, \hat{\beta}, \gamma, \hat{\delta}) \in L_{\mathsf{ua}}$.

Theorem 3.1. *Assuming the existence of collision-resistant hash functions, the protocol in Fig 3 is a 6-round public-coin non-malleable zero-knowledge argument of knowledge system for* **NP**.

Proof 1. *The definition of the language in Fig 2 is essentially the same as the definition in [PR05b],*

Relation $\mathbf{R}_{\mathsf{sim}}$:

For a triplet $\langle h, c, r\rangle \in \mathcal{H}_n \times \{0,1\}^n \times \{0,1\}^{2n}$, we say a relation $\mathbf{R}_{\mathsf{sim}}((h,c,r),(M,\rho_1,y)) = 1$, iff $M \in \{0,1\}^{n^{\omega(1)}}, \rho_1 \in \{0,1\}^{poly(n)}, |y| \leq |r| - n$; $c = \mathsf{Com}(h(M), \rho_1)$ such that $M(y) = r$ within $n^{\omega(1)}$ steps.

Language L_{ua} :

We say $(h, c, r, \hat{\beta}, \gamma, \hat{\delta}) \in L_{\mathsf{ua}}$, iff there exist $(M, \rho_1, y, \beta, \rho_2, \delta, \rho_3)$ such that

- $M \in \{0,1\}^{n^{\omega(1)}}, \rho_1, \rho_2, \rho_3 \in \{0,1\}^{poly(n)}, |y| \leq |r| - n$;
- $c = \mathsf{Com}(h(M), \rho_1)$, $\hat{\beta} = \mathsf{Com}(\beta, \rho_2)$, $\hat{\delta} = \mathsf{Com}(\delta, \rho_3)$;
- $(h, \beta, \gamma, \delta)$ is an accepting transcript of universal argument proving the statement: such that $((h,c,r),(M,\rho_1,y)) \in R_{\mathsf{sim}}$.

Figure 2: The languages used in public-coin NMZK

so the protocol in Fig 3 is also a public-coin *zero-knowledge argument. The* completeness *and* soundness *follow quite naturally.*

Completeness. *Roughly speaking, the* completeness *can be obtained from the* correctness *of NMCom and the completeness of sWI.*

Soundness. *The* soundness *of protocol can be obtained from the* binding property *of the NMCom and the* soundness *of sWI used in stage two. In particular, assume $x \notin L$, if there exists a PPT cheating P^* which can convince the verifier in stage two. From the definition of L_{ua}, there must exist an accepting universal argument transcript $(h, \beta, \gamma, \delta)$ which can prove the relation $\boldsymbol{R}_{sim}((h,c,r),(M,s,y))$ is true* except with negligible probability. *That is there exists a PPT machine M on input a short bit string y (of length bounded in n) which can predict the challenge message r (length of $2n$). However, this is* information theoretically *impossible. Thus, we reach a contradiction through violating the soundness of Barak's protocol.*

NMZK. *Now we sketch how to build the simulator-extractor to prove the non-malleable zero-knowledge. First, we construct a PPT simulator S that simulates the view of $\mathcal{A}$ but does not extract witnesses in the right session. Then, we construct a PPT simulator-extractor SE via intermediate simulator S that simulates the view of $\mathcal{A}$ and extracts the witness from the extractable non-malleable commitment NMCom.*

More specifically, S internally invokes $\mathcal{A}$ and interacts with $\mathcal{A}$ as honest prover and honest verifier in the following way. To simulate the view in the right interaction, S simply follows the honest verifier strategy. To simulate the view in the left interaction, S commits a dummy string (i.e., 0^n) by invoking NMCom twice to generate the transcripts $(\mathsf{nm}_1^1(0^n), \mathsf{nm}_2^1, \mathsf{nm}_3^1(0^n))$ and $(\mathsf{nm}_1^2(0^n), \mathsf{nm}_2^2, \mathsf{nm}_3^2(0^n))$ and uses the code description of the adversary $\mathcal{A}$ as the fake witness in a straight-line *manner to generate the transcript of the "special purpose" universal argument $(h, c, r, \hat{\beta}, \gamma, \hat{\delta})$ and $(\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3)$. We denote the code description of $\mathcal{A}$ as M and the code description of the honest verifier as V. In the first stage, upon receiving the hash function h from the adversary $\mathcal{A}$, S sends $c = com(h(M,V), \rho_1)$ and $\mathsf{nm}_1^1(0^n)$ to $\mathcal{A}$ and receives the corresponding responses r, nm_2^1. In the second stage, because S has the witness for statement (h,c,r), it can complete*

Common input: $x \in L$ and identity $\mathsf{id} \in \{0,1\}^n$.
Auxiliary input to P **:** $w \in R_L(x)$.

Stage one: P and V generate the trapdoor statement by running Barak's protocol and compute the first non-malleable commitment to the witness w under the identity id.

1. The verifier computes $h \xleftarrow{R} \mathcal{H}_n$ and sends h to P.
2. The prover computes $c = \mathsf{Com}(0^n, \rho_1)$ using Com and the first round message $\mathsf{nm}_1^1(w, s_1)$ using NMCom, and sends c, nm_1^1 to V.
3. The verifier computes $r \xleftarrow{R} \{0,1\}^{2n}$ and the second round message $\mathsf{nm}_2^1 \xleftarrow{R} \{0,1\}^n$ using NMCom, and sends r, nm_2^1 to P.

Stage two: P and V run the three round sWI, NMCom and "encrypted" sUA in parallel, where P will compute a second non-malleable commitment to the witness w under the identity id.

4. The prover computes
    - the third round message $\mathsf{nm}_3^1(w, s_2)$ to complete the first non-malleable commitment and the first round message $\mathsf{nm}_1^2(w, s_3)$ using NMCom under the identity id,
    - $\hat{\beta} = \mathsf{Com}(0^n, \rho_2)$ using Com,
    - wi_1 using sWI,

    and sends $(\hat{\beta}, \mathsf{wi}_1, \mathsf{nm}_3^1, \mathsf{nm}_1^2)$.
5. The verifier computes $\mathsf{nm}_1^2 \xleftarrow{R} \{0,1\}^n, \gamma \xleftarrow{R} \{0,1\}^n, \mathsf{wi}_2 \xleftarrow{R} \{0,1\}^n$ and sends $(\gamma, \mathsf{wi}_2, \mathsf{nm}_1^2)$.
6. The prover computes
    - the third round message $\mathsf{nm}_3^2(w, s_4)$ to complete the second non-malleable commitment,
    - $\hat{\delta} = \mathsf{Com}(0^n, \rho_3)$ using Com and wi_3 using sWI,

    and sends $(\hat{\delta}, \mathsf{wi}_3, \mathsf{nm}_3^2)$.

V accepts if the transcript $(\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3)$ is an accepting proof for the following statements:

- $\exists(w, \mathsf{dec}_1, \mathsf{dec}_2)$ s.t. $R_L(x, w) = 1$ and (w, dec_1) is the decommitment for the transcript $(\mathsf{id}, \mathsf{nm}_1^1, \mathsf{nm}_2^1, \mathsf{nm}_3^1)$ and (w, dec_2) is the decommitment for the transcript $(\mathsf{id}, \mathsf{nm}_1^2, \mathsf{nm}_2^2, \mathsf{nm}_3^2)$ **or**
- $\exists(M, \rho_1, y, \beta, \rho_2, \delta, \rho_3)$ s.t. $(h, c, r, \hat{\beta}, \gamma, \hat{\delta}) \in L_{\mathsf{ua}}$.

Figure 3: 6-round public-coin non-malleable zero-knowledge argument

the encrypted universal argument proof and special WI as follows:

- *The simulator S completes the first non-malleable commitment by computing* $\mathsf{nm}_3^1(0^n)$ *and executes the second non-malleable commitment by computing* $\mathsf{nm}_1^2(0^n)$*. Next, S invokes the underlying PCP system to generate the PCP proof and then constructs the Markle tree HT of this proof under the hash function* h*. Denote the root of the Markle tree as* β*, and then it computes* $\hat{\beta} = Com(\beta, \rho_1)$*. Finally, S acts as an honest prover to compute* wi_1 *and sends* $(\hat{\beta}, \mathsf{wi}_1, \mathsf{nm}_3^1, \mathsf{nm}_1^2)$ *to* $\mathcal{A}$*.*

- *Upon receiving the fifth message* $(\gamma, \mathsf{wi}_2, \mathsf{nm}^2_1)$ *from* $\mathcal{A}$*,* S *invokes* PCP *system on input* $((h, c, r), \gamma)$ *to generate* PCP *queries* Q*, and computes* PCP *answers* $\delta = \{q, \sigma_q, \mathsf{auth}_q(\sigma)\}_{q \in Q}$*, where* σ_q *is the* q*-th bit of* σ *and* $\mathsf{auth}_q(\sigma)$ *is the certificate path of* HT *for* σ_q*. Then it computes* $\hat{\delta} = Com(\delta, \rho_2)$*. Next,* S *completes the second non-malleable commitment by computing* $\mathsf{nm}^2_3(0^n)$ *and computes* wi_3 *such that* $(\mathsf{wi}_1, \mathsf{wi}_2, \mathsf{wi}_3)$ *is a proof for the statement* $(h, c, r, \hat{\beta}, \gamma, \hat{\delta}) \in L_{\mathsf{ua}}$*. Now,* S *sends* $(\hat{\delta}, \mathsf{wi}_3, \mathsf{nm}^2_3)$ *to* $\mathcal{A}$*.*

Finally, the simulator S *outputs the view of the adversary* $\mathcal{A}$*. We denote the simulated view as* $\{\mathsf{sim\text{-}view}_{\mathcal{A}}(1^n, x, z)\}_{n \in N, x \in L \cap \{0,1\}^n, z \in \{0,1\}^*}$ *and the real view as* $\{\mathsf{real\text{-}view}_{\mathcal{A}}(1^n, x, z)\}_{n \in N, x \in L \cap \{0,1\}^n, z \in \{0,1\}^*}$*.*

Analysis of the Simulator. *From the language* L_{ua} *defined in Fig 2, we know that if the simulator can* non-black box *access the adversary* $\mathcal{A}$*, then it uses the* code description *of the adversary* $\mathcal{A}$ *as witness can complete the proof for the statement in* L_{ua} *in stage two. Thus, the* correctness *of* S *can be directly obtained from the* completeness *of* sUA*,* NMCom *and* sWI*.*

The computational indistinguishability *of* $\{\mathsf{real\text{-}view}_{\mathcal{A}}\}$ *and* $\{\mathsf{sim\text{-}view}_{\mathcal{A}}\}$ *can be obtained from the* computational-hiding *of the* Com *and* NMCom *and the* witness-indistinguishability *of the* sWI*. Roughly speaking, we can consider the following hybrid experiments. Let* Hyb_0 *denotes the real experiment which outputs the real view* $\{\mathsf{real\text{-}view}_{\mathcal{A}}\}$ *and* Hyb_3 *denotes the simulated experiment by the simulator* S *which outputs the simulated view* $\{\mathsf{sim\text{-}view}_{\mathcal{A}}\}$*.*

We define Hyb_1 *in the same way as the* Hyb_0 *except that the simulator* S *use both the witness* w *and the adversary's code* M *to complete the protocol execution. More specially, in* Hyb_1*,* S *uses the fake witness* M *to complete the "special-purpose" universal argument in the left interaction where* $c = \mathsf{Com}(h(M, V), \rho_1)$*,* $\hat{\beta} = \mathsf{Com}(\beta, \rho_2)$ *and* $\hat{\delta} = \mathsf{Com}(\delta, \rho_3)$*. However, in stage two* S *still prove the OR statement using the witness* w *and* dec_1 *and* dec_2*.*

Thus, the computational indistinguishability *of the output of* Hyb_0 *and* Hyb_1 *can be followed from the* computational-hiding *of the* Com*.*

We define Hyb_2 *in the same way as the* Hyb_1 *except that in stage two* S *prove the OR statement using the witness* $(h, (M, V), \rho_1, c, \beta, \rho_2, \delta, \rho_3)$*. Thus, the* computational indistinguishability *of the output of* Hyb_1 *and* Hyb_2 *can be followed from the* witness-indistinguishability *of the* sWI*.*

The only difference between the output of Hyb_2 *and* Hyb_3 *is that, in the former the transcripts* $(\mathsf{nm}^1_1, \mathsf{nm}^1_2, \mathsf{nm}^1_3)$ *and* $(\mathsf{nm}^2_1, \mathsf{nm}^2_2, \mathsf{nm}^2_3)$ *are generated using the value* w *, and in the latter, the transcripts* $(\mathsf{nm}^1_1, \mathsf{nm}^1_2, \mathsf{nm}^1_3)$ *and* $(\mathsf{nm}^1_2, \mathsf{nm}^2_2, \mathsf{nm}^2_3)$ *are generated using* 0^n*.*

Thus, the computational indistinguishability *of the output of* Hyb_2 *and* Hyb_3 *can be followed from the* computational-hiding *of the* NMCom*.*

Combining the above, we can argue that the output of Hyb_0 *and* Hyb_3 *are* computationally indistinguishable*, that is* $\{\mathsf{real\text{-}view}_{\mathcal{A}}\} \overset{c}{\approx} \{\mathsf{sim\text{-}view}_{\mathcal{A}}\}$*.*

Simulator-Extractor SE***.*** *We use* SE *to simulate the view of* $\mathcal{A}$ *by executing* S *as the first part of its output. Now considering the right interaction is accepted and* $\widetilde{\mathsf{id}}$ *is different from* id *in the left interaction, we will show that the extracted witness is indeed the witnesses of the statement proved in the right interaction.*

Observe that in the experiment Hyb_0*, the simulator* S *holds the real witnesses of the left interaction and just acts as an honest prover in the interaction and an honest verifier in the right interaction. Then following from the soundness of the* WIPOK *and the honest-extractable property of the* NMCom*, we can conclude that for any accepting right interaction and the right* $\widetilde{\mathsf{id}}$ *different from the left* id*,* $\mathcal{A}$ *commits successfully a real witness in the* NMCom except with negligible probability*. That is* $\mathcal{A}$ *never cheats in* Hyb_0*, and the simulator-extractor* SE *can extract the witness by rewinding the first (*or the second*) non-malleable commitment from the third round to the second round* *(*or the sixth round to the fifth round*)* except with negligible probability*. In order to prove the non-malleable zero-knowledge, we need to prove that $\mathcal{A}$ never cheats in Hyb_1 and Hyb_2 also, which means that $\mathcal{A}$ will commit the same witness $\widetilde{w}$ in the non-malleable commitments on the right.*

Recall that the adversary $\mathcal{A}$ controls the message executing in the two sides and there are many adversarial schedules. According to when the adversary $\mathcal{A}$ sends the fourth round message on the right, we divide the schedules into two cases and prove them separately.

***Schedule 1:** The adversary $\mathcal{A}$ sends the fourth round message on the right before it receives the sixth round message on the left. In such schedule, we observe that the fifth round messages on the left are sent by the adversary $\mathcal{A}$ to the prover. So we only need to focus on the first four rounds messages of the left which may interrupt the right side.*

Recall that the only difference of the message received in the first four rounds of the left side between Hyb_0 and Hyb_1 is that, in the former, the commitments $c = \mathsf{Com}(0^n, \rho_1), \hat{\beta} = \mathsf{Com}(0^n, \rho_2)$, in the latter the commitments $c = \mathsf{Com}(h(M, V), \rho_1'), \hat{\beta} = \mathsf{Com}(\beta, \rho_2')$. We know that any non-malleable commitment scheme is non-malleable w.r.t to any non-interactive primitives, which means that the right rewinding on the first non-malleable commitment does not interrupt the security of the non-interactive computational hiding *commitment scheme Com. Thus, we can prove that $\mathcal{A}$ never cheats in Hyb_1* except with negligible probability*, otherwise we can break the* computational hiding property *of Com.*

Next, the only difference of the message received in the first four rounds of the left side between Hyb_1 and Hyb_2 is the message wi_1. From the delay-input sWI, we know that wi_1 are just the commitments generated by Com which is independent of the statement to be proved. Thus, we can prove that $\mathcal{A}$ never cheats in Hyb_2 except with negligible probability*, otherwise we can break the* computational hiding property *of Com.*

Next, recall that the only difference of the message received in the first four rounds of the left sides between Hyb_2 and Hyb_3 is that, in the former, the messages $\mathsf{nm}_1^1, \mathsf{nm}_3^1, \mathsf{nm}_1^2$ are generated using the value w, and in the latter the messages $\mathsf{nm}_1^1, \mathsf{nm}_3^1, \mathsf{nm}_1^2$ are generated using the value 0^n. Here, we further consider the two types of schedules.

- *The first type is that the third and the fourth round messages (which contains $\widetilde{\mathsf{nm}}_2^1$ and $\widetilde{\mathsf{nm}}_3^1$) on the right cover the third and fourth round messages (which contains nm_2^1 and nm_3^1) on the left but not cover the second round messages (which contains nm_1^1) on the left. In such condition, the advantage of the adversary $\mathcal{A}$ is equal to the advantage of the adversary $\mathcal{A}$ who execute the right four rounds messages* in parallel *with the left four rounds messages. In such case we can see the adversary $\mathcal{A}$ as a* synchronous adversary*. Because the* three-round non-malleable commitment *we use is a* non-malleable *against* a synchronizing adversary *(see [GPR16]), we can argue that if in Hyb_2 the adversary commits the values $\widetilde{w}$ in the transcript $(\widetilde{\mathsf{nm}}_1^1, \widetilde{\mathsf{nm}}_2^1, \widetilde{\mathsf{nm}}_3^1)$, then in Hyb_3 the adversary will also commit the same value $\widetilde{w}$ in the transcript $(\widetilde{\mathsf{nm}}_1^1, \widetilde{\mathsf{nm}}_2^1, \widetilde{\mathsf{nm}}_3^1)$* except with negligible probability*, otherwise we can break the* non-malleable property *of the NMCom (we note that this idea was first used in [COSV17a]).*

- *The second type includes the schedules other than the above. That is the* right four rounds messages *are executed not in parallel with the* left four rounds messages*. In such case, we can reduce the security to the* computational-hiding *of the non-malleable commitment scheme. Because when we rewind the right three round non-malleable commitment to extract the commitment value $\widetilde{w}$, the left three round non-malleable commitment will not be rewound and its* computational-hiding *is preserved. If the adversary $\mathcal{A}$ cheats in Hyb_3, that is in Hyb_2 and Hyb_3 the commitment value which we extract are different with* noticeable probability*, then we can obtain a contradiction.*

Schedule 2: *The adversary $\mathcal{A}$ sends the* fourth round message *on the right after it received the* sixth round message *on the left. This case can be easily proved. Because in such condition, the* second non-malleable commitment *is fully executed after the execution of the left protocol. As before, we can reduce the security to the* computational-hiding *of the non-malleable commitment scheme NMCom and the* non-interactive *commitment Com and the* witness-indistinguishability *of the sWI. More specifically, for Hyb_0 and Hyb_1, if the adversary $\mathcal{A}$ cheats in Hyb_1, then we can break the* computational-hiding *of the non-interactive commitment Com. For Hyb_1 and Hyb_2, if the adversary $\mathcal{A}$ cheats in Hyb_2, then we can break the* witness-indistinguishability *of the sWI. For Hyb_2 and Hyb_3, if the adversary $\mathcal{A}$ cheats in Hyb_3, then we can break the* computational-hiding *of the non-malleable commitment scheme NMCom.*

Put the three above together, we obtain that the extractor does not break the security of the left protocol no matter for the simulator or for the honest prover on the left. Because for the simulator S we have that $\{\mathsf{sim\text{-}view}_{\mathcal{A}}(1^n, x, z)\} \overset{c}{\approx} \{\mathsf{real\text{-}view}_{\mathcal{A}}(1^n, x, z)\}$, thus we can conclude that for the simulator-extractor SE, it holds that $\{\mathsf{sim\text{-}view}_{\mathcal{A}}(1^n, x, z), \widetilde{w}\} \overset{c}{\approx} \{\mathsf{real\text{-}view}_{\mathcal{A}}(1^n, x, z), \widetilde{w}\}$ for any right interaction that is accepting and uses a different identity from the left interaction.

Combining the above analysis together, we complete the proof of public-coin non-malleable zero-knowledge property. □

Remark 1. *We remark that the protocol and its soundness proof described above relies on* collision resistant hash functions *against* slightly super-polynomial *adversaries. In order to only rely on* collision resistance *against polynomially adversaries, we should use the "error-correcting code" ECC (i.e., with constant distance and with polynomial-time encoding and decoding) technique first appeared in [BG08]. More specifically, we replace the commitment of the form $\mathsf{Com}(h(M))$ with $\mathsf{Com}(h(\mathsf{ECC}(M)))$ in stage 1. The language relation R_{sim} we define in Fig. 2 should modify as follows: $\boldsymbol{R}_{\mathsf{sim}}((h, c, r), (M, s, y)) = 1$ iff $c = \mathsf{Com}(|\mathsf{ECC}(M)|, h(\mathsf{ECC}(M)))$ and $M(y) = r$, where $h(\mathsf{ECC}(M))$ is the root of the markle tree of the $\mathsf{ECC}(M)$ under the hash function h.*

3.2 3-round Non-malleable Zero-Knowledge

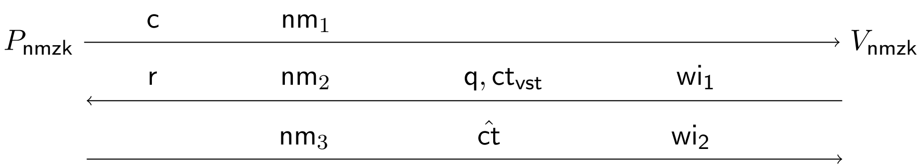
In this section we give our construction of the *3-round non-malleable zero-knowledge argument* protocol. We use the following building blocks:

- Non-interactive *perfectly-binding commitment scheme*: Com.
- 3-round *public-coin honest-extractable non-malleable commitment scheme*: NMCom.
- 2-round *delayed-input statistical witness indistinguishable arguments* : SWI.
- 2-round weak memory delegation scheme for quasi-polynomial bounded computations: MD.
- Semantically secure and circuit-private, 1-hop homomorphic encryption scheme: (Enc, Eval, Dec).

For our propose, we introduce the following notation [BKP18, BBK+16] which was used to achieve the variant of the *two-message memory delegation technique* [CKLR11] described in the introduction.

- For a well-formed string $\mathsf{mes} = (\mathsf{c}, \mathsf{nm}_1)$, denote by M_{mes} a Turing machine as follows: On input the memory $D = V^*$, M_{mes} parses V^* as a Turing machine, runs V^* on input mes, parses the result as $(\mathsf{r}, \mathsf{nm}_2, \mathsf{q}, \mathsf{ct}_{\mathsf{vst}}, \mathsf{wi}_1)$, and outputs r.

- For a well-formed string $\mathsf{param} = (\mathsf{mes}, \mathsf{d}, \mathsf{t}, \mathsf{r}, \pi, \mathsf{q})$, denote by $\mathbb{C}_{\mathsf{param}}$ a circuit with param hard-coded as follows: on input a verification state vst, $\mathbb{C}_{\mathsf{param}}(\mathsf{vst})$ outputs
    - 1 if $\mathsf{MD.Ver}(\mathsf{d}, \mathsf{M}_{\mathsf{mes}}, \mathsf{t}, \mathsf{r}, \mathsf{vst}, \pi) = 1$ or $(\mathsf{q}, \mathsf{vst}) \neq \mathsf{MD.Query}(1^n)$.
    - 0 otherwise.
- Denote by $\mathbb{C}_1$ a circuit of the same size as the circuit $\mathbb{C}_{\mathsf{param}}$ that always returns 1.



- (c, r) is the transcript of the "encrypted" two-message memory delegation, where c is a commitment to $(0^n, 0^{\log n})$, and r is the challenge message.
- $(\mathsf{nm}_1, \mathsf{nm}_2, \mathsf{nm}_3)$ is the transcript of the commitment to the witness w computed using the honest-extractable non-malleable commitment NMCom under the identity id.
- $(\mathsf{wi}_1, \mathsf{wi}_2)$ is the transcript of the SWI proving the knowledge of the opening of the non-malleable commitment to w such that $R(x, w) \in R_L$ **or** the knowledge of $(\mathsf{d}, \mathsf{t}, \rho, \pi)$ such that $\mathsf{c} = \mathsf{Com}(\mathsf{d}, \mathsf{t}; \rho)$ and $\hat{\mathsf{ct}} = \mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}})$.

Figure 4: 3-round non-malleable zero-knowledge for NP

Theorem 3.2. *Assuming keyless multi-collision resistant hash functions, LWE and DDH (or QR or N^{th} residuosity), all quasi-polynomially hard, there exist 3-round non-malleable zero-knowledge arguments for* ***NP****.*

Proof 2. *The* completeness *can be obtained from the* correctness *of NMCom and the* completeness *of SWI. The* soundness *and* zero-knowledge *essentially can be proved in the same way as the protocol in [BKP18]. For completeness, we briefly describe the proof, and we refer the reader to [BKP18] for more detail about this part.*

Soundness. *Assuming that $x \notin L$, in order to pass the SWI with respect to an evaluated cipher $\hat{\mathsf{ct}}$ that decrypts to 1, the prover must know a digest d, a time bound t, and proof π, such that $\mathbb{C}_{param}(vst) = 1$. This, by definition, means that $(\mathsf{d}, \mathsf{t}, \pi)$ are such that the delegation verifier Ver is convinced that the digest d corresponds to a machine V^* such that $V^*(\mathsf{c}, \mathsf{nm}_1) = \mathsf{r}$. Intuitively, this implies that the prover manages to commit to a program that predicts the random string r before it was ever sent, which is unlikely. Specifically, we can construct an attacker from this prover which can sample a digest d and a computation M such that* with noticeable probability *ϵ over a random output $r \in \{0, 1\}^n$, it successfully produces a convincing proof consistently with(M, r), which is enough to break the* weak soundness *of memory delegation in Definition 2.7. Here we will also rely on the* semantic security *of the encryption scheme to claim that the encrypted verification state $\mathsf{ct}_{\mathsf{vst}}$ is* hiding.

NMZK. *Now we sketch how to build the simulator-extractor to prove the non-malleable zero-knowledge. First, we construct a PPT simulator S that simulates the view of $\mathcal{A}$ but does not extract*

Common input: $x \in L$ and identity $\mathsf{id} \in \{0,1\}^n$.
Auxiliary input to P : $w \in R_L(x)$.

P and V run the three round SWI, NMCom and "encrypted" MD in parallel.

1. The prover computes
   - $c = \mathsf{Com}(0^n, 0^{\log \gamma(n)}; \rho)$, where $\gamma = n^{\log\log(n)}$,
   - the first round message $\mathsf{nm}_1(w, s_1)$ using NMCom,

   and sends $(\mathsf{c}, \mathsf{nm}_1)$ to V.
2. The verifier computes
   - $r \xleftarrow{R} \{0,1\}^n$,
   - the second round message $\mathsf{nm}_2 \xleftarrow{R} \{0,1\}^n$ using NMCom,
   - $(\mathsf{q}, \mathsf{vst}) \leftarrow \mathsf{MD.Query}(1^n)$, where we assume w.l.o.g that vst consists of the coins of MD.Query,
   - $(\mathsf{ct}_{\mathsf{vst}}, \mathsf{sk}) \leftarrow \mathsf{Enc}_{\mathsf{sk}}(\mathsf{vst})$, an encryption of the verification state,
   - the first round message wi_1 using the SWI,

   and sends $(\mathsf{r}, \mathsf{nm}_2, \mathsf{q}, \mathsf{ct}_{\mathsf{vst}}, \mathsf{wi}_1)$ to P.
3. The prover computes
   - the third round message $\mathsf{nm}_3(w, s_2)$ to complete the non-malleable commitment,
   - $\hat{\mathsf{ct}} \leftarrow \mathsf{Eval}(\mathbb{C}_1, \mathsf{ct}_{\mathsf{vst}})$, an evaluation of the constant one function,
   - the second round message wi_2 using the SWI,

   and sends $(\mathsf{nm}_3, \hat{\mathsf{ct}}, \mathsf{wi}_2)$ to V.

V accepts if $\mathsf{Dec}_{\mathsf{sk}}(\hat{\mathsf{ct}}) = 1$ and the transcript $(\mathsf{wi}_1, \mathsf{wi}_2)$ is an accepting proof for the following statements:

- $\exists(w, \mathsf{dec} \in \{0,1\}^{\mathsf{poly(n)}})$ s.t. $R_L(x, w) = 1$ and (w, dec) is the decommitment for the transcript $(\mathsf{id}, \mathsf{nm}_1, \mathsf{nm}_2, \mathsf{nm}_3)$ **or**
- $\exists(\mathsf{d}, \pi, \rho \in \{0,1\}^{\mathsf{n}}, \mathsf{t} < \gamma(\mathsf{n}))$ s.t. $\mathsf{c} = \mathsf{Com}(\mathsf{d}, \mathsf{t}; \rho)$ and $\hat{\mathsf{ct}} = \mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}})$ where $\mathsf{param} = (\mathsf{mes} = (\mathsf{c}, \mathsf{nm}_1), \mathsf{d}, \mathsf{t}, \mathsf{r}, \pi, \mathsf{q})$.

Figure 5: 3-round non-malleable zero-knowledge argument of knowledge for NP

witnesses in the right session. Then, we construct a PPT simulator-extractor SE via intermediate simulator S that simulates the view of $\mathcal{A}$ and extracts the witnesses from the extractable non-malleable commitment NMCom.

More specifically, S internally invokes $\mathcal{A}$ and interacts with $\mathcal{A}$ as honest prover and honest verifier in the following way. To simulate the view in the right interaction, S simply follows the honest verifier strategy. To simulate the view in the left interaction, S commits a dummy string (i.e., 0^n) by invoking NMCom to generate the transcripts $(nm_1(0^n), nm_2, nm_3(0^n))$, and uses the code description of the adversary $\mathcal{A}$ as the fake witness in a straight-line *manner to generate the transcript of the* $(\mathsf{c}, \mathsf{r}, \mathsf{q}, \mathsf{ct}_{\mathsf{vst}}, \hat{\mathsf{ct}})$ *and* (wi_1, wi_2).

More specifically, for a statement $x \in L$ where $|x| = n$, we denote the code description of $\mathcal{A}$

as M and the code description of the honest verifier as V, a polynomial bound $t(n) = n^{O(1)}$ on its running time. The simulator S operates on the left side as follows:

- *In the first round, S computes the digest* $\mathsf{d} = \mathsf{MD.Mem}(1^{\mathsf{n}}, (\mathsf{M}, \mathsf{V}))$ *and then computes a commitment* $\mathsf{c} = \mathsf{Com}(\mathsf{d}, \mathsf{t}; \rho)$ *to the digest d and M's running time* t *using random coins $\rho \xleftarrow{R} \{0,1\}^n$. Next, it computes the first message of the NMCom* $\mathsf{nm}_1 = \mathsf{nm}_1(0^n)$ *to a dummy string 0^n and sends* $(\mathsf{c}, \mathsf{nm}_1)$ *to $\mathcal{A}$. Here, we set* $\mathsf{mes} = (\mathsf{c}, \mathsf{nm}_1)$.

- *In the third round, upon receiving the responses* $(\mathsf{r}, \mathsf{nm}_2, \mathsf{q}, \mathsf{ct}_{\mathsf{vst}}, \mathsf{wi}_1)$ *from $\mathcal{A}$, S computes the proof* $\pi = \mathsf{MD.Prov}(1^{\mathsf{n}}, (\mathit{M}, \mathit{V}), (\mathsf{M}_{\mathsf{mes}}, \mathsf{t}, \mathsf{r}), \mathsf{q})$ *for the memory computation that* $\mathsf{M}_{\mathsf{mes}}$ *on input the memory (M, V) output r within t steps. Here, we set* $\mathsf{param} = (\mathsf{mes} = (\mathsf{c}, \mathsf{nm}_1), \mathsf{d}, \mathsf{t}, \mathsf{r}, \pi, \mathsf{q})$. *Now, S computes* $\hat{\mathsf{ct}} \leftarrow \mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}})$ *and the third message of the NMCom* $\mathsf{nm}_3 = \mathsf{nm}_3(0^n)$ *to the dummy string 0^n. Next, S computes the second SWI message* wi_2 *for the statement that it knows the opening of the non-malleable commitment to w such that $R(x, w) \in R_L$* ***or*** *$c = \mathit{Com}(d, t; \rho)$ and $\hat{ct} = \mathit{Eval}(\mathbb{C}_{\mathit{param}}, \mathit{ct}_{\mathit{vst}})$ using the witness* $(\mathsf{d}, \mathsf{t}, \rho, \pi)$. *Then, S sends* $(\mathsf{nm}_3, \hat{\mathsf{ct}}, \mathsf{wi}_2)$ *to $\mathcal{A}$.*

Finally, the simulator S outputs the view of the adversary $\mathcal{A}$. We denote the simulated view as $\{\textit{sim-view}_{\mathcal{A}}(1^n, x, z)\}_{n \in N, x \in L \cap \{0,1\}^n, z \in \{0,1\}^*}$ and the real view as $\{\textit{real-view}_{\mathcal{A}}(1^n, x, z)\}_{n \in N, x \in L \cap \{0,1\}^n, z \in \{0,1\}^*}$.

Analysis of the Simulator. *From the definition of the Turing machine* $\mathsf{M}_{\mathsf{mes}}$ *and the circuit $\mathbb{C}_{\mathit{param}}$, if the simulator can non-black box access the adversary $\mathcal{A}$, then it uses the code description of the adversary $\mathcal{A}$ to predict the challenge string r* except with negligible probability. *Therefore, the* correctness *of S can be directly obtained from the* correctness *of MD, the* perfect correctness *of the 1-hop homomorphic encryption scheme and the* completeness *of the SWI.*

The computational indistinguishability *of $\{\textit{real-view}_{\mathcal{A}}\}$ and $\{\textit{sim-view}_{\mathcal{A}}\}$ can be obtained from the* computational-hiding *of the Com and NMCom and the* witness-indistinguishability *of the SWI. Roughly speaking, we consider a simulator S that it uses both the witness w and the adversary's code M to complete the protocol. Now, we define the hybrid experiments required to prove the zero-knowledge.*

Let Hyb_0 denotes the real experiment which outputs the real view $\{\textit{real-view}_{\mathcal{A}}\}$. In particular, for $x \in L$ and $(x, w) \in R_L$, the simulator acts as an honest prover who uses the witness w on the left to interactive with the adversary $\mathcal{A}$, and acts as an honest verifier in the right to interactive with $\mathcal{A}$.

The hybrid Hyb_1 is identical to Hyb_0 except that the simulator S uses the trapdoor M to compute the commitment $\mathsf{c} = \mathsf{Com}(\mathsf{d}, \mathsf{t}; \rho)$ *to the digest d and M's running time t. It is easy to see that the* computational indistinguishability *of the output of Hyb_0 and Hyb_1 can be followed from the* computational-hiding *of the Com.*

The hybrid Hyb_2 is identical to Hyb_1 except that the simulator S computes $\hat{\mathsf{ct}}$ *by invoking* $\mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}})$ *instead of* $\mathsf{Eval}(\mathbb{C}_1, \mathsf{ct}_{\mathsf{vst}})$. *The* computational indistinguishability *of the output of Hyb_2 and Hyb_1 can be followed from the* circuit privacy *of the 1-hop homomorphic encryption. More specifically, if the query is inconsistent with the query q, i.e.,* $(\mathsf{q}, \mathsf{vst}^*) \neq \mathsf{MD.Query}(1^{\mathsf{n}})$, *then by the definition in the beginning it holds that* $\mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{vst}^*) = 1$. *Thus, for an illegal encryption ct^*, by the* circuit privacy, *we have that* $\mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}^*) \overset{c}{\approx} \mathsf{S}_{\mathsf{1hop}}(\mathsf{ct}^*, \perp, |\mathbb{C}_{\mathsf{param}}|) \equiv \mathsf{S}_{\mathsf{1hop}}(\mathsf{ct}^*, \perp, |\mathbb{C}_1|) \overset{c}{\approx} \mathsf{Eval}(\mathbb{C}_1, \mathsf{ct}^*)$. *Otherwise, for any vst which is consistent with the query q, by the* perfect completeness *of the delegation scheme it holds that* $\mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{vst}) = 1$. *Also, by the* circuit privacy, *we have that* $\mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}}) \overset{c}{\approx} \mathsf{S}_{\mathsf{1hop}}(\mathsf{ct}_{\mathsf{vst}}, \mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{vst}), |\mathbb{C}_{\mathsf{param}}|) \equiv \mathsf{S}_{\mathsf{1hop}}(\mathsf{ct}_{\mathsf{vst}}, \mathbb{C}_1, |\mathbb{C}_1|) \overset{c}{\approx} \mathsf{Eval}(\mathbb{C}_1, \mathsf{ct}_{\mathsf{vst}})$.

The hybrid Hyb_3 is identical to Hyb_2 except that the simulator S uses the witness $(\mathsf{d}, \mathsf{t}, \rho, \pi)$ instead of w to complete the SWI on the left. It is easy to see that this hybrid is statistically witness indistinguishable from Hyb_2.

The hybrid Hyb_4 is identical to Hyb_3 except that the simulator S computes the non-malleable commitment $(\mathsf{nm}_1, \mathsf{nm}_2, \mathsf{nm}_3)$ to a dummy string 0^n instead of the witness w. Thus, the computational indistinguishability of the output of Hyb_4 and Hyb_3 can be followed from the computational-hiding of the NMCom.

Combining the above, we can argue that the output of Hyb_0 and Hyb_4 are computationally indistinguishable. The hybrid experiment Hyb_4 is identical to with the simulator S we describe above, therefore we get $\{\textit{real-view}_{\mathcal{A}}\} \overset{c}{\approx} \{\textit{sim-view}_{\mathcal{A}}\}$.

Simulator-Extractor SE. We use SE to simulate the view of $\mathcal{A}$ by executing S as the first part of its output. Now considering the right interaction is accepted and $\widetilde{\mathsf{id}}$ is different from id in the left interaction, we will show that the extracted witness is indeed the $\boldsymbol{NP}$ witnesses of the statement proved in the right interaction.

Observe that in the experiment Hyb_0, the simulator S holds the real witnesses of the left interaction and just acts as an honest prover in the interaction and an honest verifier in the right interaction. Then following from the soundness of the SWI and the honest-extractable property of the NMCom, we can conclude that for any accepting right interaction and the right $\widetilde{\mathsf{id}}$ different from the left id, $\mathcal{A}$ commits successfully a real witness in the NMCom except with negligible probability. That is $\mathcal{A}$ never cheats in Hyb_0 and the simulator-extractor SE can extract the witness by rewinding the non-malleable commitment from the third round to the second round except with negligible probability. In order to prove the non-malleable zero-knowledge, we need to prove that $\mathcal{A}$ never cheats in the hybrid experiments from Hyb_1 to Hyb_4, which means that $\mathcal{A}$ will commit the same witness $\widetilde{w}$ in the non-malleable commitments on the right.

Recall that the adversary $\mathcal{A}$ controls the message scheduling in the two sides due to the left and right protocol are both three rounds, so there are only two different type of adversarial schedules:

Schedule 1: The left protocol and the right protocol are executed in parallel. In this schedule, we can see $\mathcal{A}$ as a synchronous adversary. We now give a series of hybrid proofs to argue the non-malleable property.

- The only difference of the message received on the left side between Hyb_0 and Hyb_1 is that, in the former $\mathsf{c} = \mathsf{Com}(0^n, 0^{\log n}; \rho)$ and in the latter $\mathsf{c} = \mathsf{Com}(\mathsf{d}, \mathsf{t}; \rho')$. We know that any non-malleable commitment scheme is non-malleable w.r.t to any non-interactive primitives, which means that the right rewinding on the non-malleable commitment does not interrupt the security of the non-interactive computational hiding commitment scheme Com. Thus, we can prove that $\mathcal{A}$ never cheats in Hyb_1 except with negligible probability, otherwise we can break the computational hiding property of Com.

- The only difference of the message received on the left side between Hyb_1 and Hyb_2 is that, in the former $\hat{\mathsf{ct}} = \mathsf{Eval}(\mathbb{C}_1, \mathsf{ct}_{\mathsf{vst}})$ and in the latter $\hat{\mathsf{ct}} = \mathsf{Eval}(\mathbb{C}_{\mathsf{param}}, \mathsf{ct}_{\mathsf{vst}})$. The same reason as before, we can argue that $\mathcal{A}$ never cheats in Hyb_2 except with negligible probability from the circuit privacy of the non-interactive 1-hop homomorphic encryption.

- The only difference of the message received on the left side between Hyb_2 and Hyb_3 is that, in the former the adversary use the witness w in SWI, and in the latter the adversary use the witness $(\mathsf{d}, \mathsf{t}, \rho, \pi)$ in SWI instead. Since SWI is statistically witness indistinguishable, we get that in Hyb_3 the adversary will also commit the same value $\widetilde{w}$ in $(\widetilde{\mathsf{nm}}_1, \widetilde{\mathsf{nm}}_2, \widetilde{\mathsf{nm}}_3))$ except with negligible probability.

- *The only difference of the message received on the left side between* Hyb_3 *and* Hyb_4 *is that, in the former the adversary commits the values* w *in the transcript* $(\mathsf{nm}_1, \mathsf{nm}_2, \mathsf{nm}_3)$*, and in the latter the adversary commits the values* 0^n *instead. Because the three-round non-malleable commitment we use is* non-malleable *against a* synchronizing adversary*, we can argue that if in* Hyb_3 *the adversary commits the values* $\widetilde{w}$ *in the transcript* $(\widetilde{\mathsf{nm}}_1, \widetilde{\mathsf{nm}}_2, \widetilde{\mathsf{nm}}_3))$*, then in* Hyb_4 *the adversary will also commit the same value* $\widetilde{w}$ *in* $(\widetilde{\mathsf{nm}}_1, \widetilde{\mathsf{nm}}_2, \widetilde{\mathsf{nm}}_3))$ except with negligible probability*, otherwise we can break the* non-malleable property *of the NMCom.*

Schedule 2: *The adversary* $\mathcal{A}$ *sends the first round message on the right after it receives the third round message on the left. In such condition, the right protocol is fully executed after the execution of the left protocol, and the simulator-extractor can extract the right witness* $\widetilde{w}$ *using rewinding approach without interfering the left execution. Therefore, we can argue the adversary* A *never cheats in* Hyb_1 *by reducing the security to the* computational-hiding *of NMCom, never cheats in* Hyb_2 *by reducing the security to the* circuit privacy *of the* non-interactive *1-hop homomorphic encryption, never cheats in* Hyb_3 *by reducing the security to the* witness-indistinguishability *of SWI,* except with negligible probability *and never cheats in* Hyb_4 *by reducing the security to the* computational-hiding *of Com.*

Put the above together, we obtain that the simulator-extractor does not break the security of the left protocol no matter for the simulator or for the honest prover on the left. Because for the simulator S we have $\{\mathsf{sim\text{-}view}_{\mathcal{A}}(1^n, x, z)\} \overset{c}{\approx} \{\mathsf{real\text{-}view}_{\mathcal{A}}(1^n, x, z)\}$*, thus we can conclude that for the simulator-extractor SE it holds that* $\{\mathsf{sim\text{-}view}_{\mathcal{A}}(1^n, x, z), \widetilde{w}\} \overset{c}{\approx} \{\mathsf{real\text{-}view}_{\mathcal{A}}(1^n, x, z), \widetilde{w}\}$ *for any right interaction that is accepting and uses a different identity from the left interaction.*

Combining the above analysis together, we complete the proof of three-round non-malleable zero-knowledge property*.* ☐

4 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61772521,No 61772522,No 61932019), Key Research Program of Frontier Sciences, CAS (QYZDB-SSW-SYS035), and the Open Project Program of the State Key Laboratory of Cryptology.

References

[Bar01] Boaz Barak. How to go beyond the black-box simulation barrier. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 106–115, 2001.

[BBK+16] Nir Bitansky, Zvika Brakerski, Yael Tauman Kalai, Omer Paneth, and Vinod Vaikuntanathan. 3-message zero knowledge against human ignorance. In *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I*, pages 57–83, 2016.

[BCPR16] Nir Bitansky, Ran Canetti, Omer Paneth, and Alon Rosen. On the existence of extractable one-way functions. *SIAM J. Comput.*, 45(5):1910–1952, 2016.

[BDRV18] Itay Berman, Akshay Degwekar, Ron D. Rothblum, and Prashant Nalini Vasudevan. Multi-collision resistant hash functions and their applications. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory* *and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings*, pages 133–161, 2018.

[BG92] Mihir Bellare and Oded Goldreich. On defining proofs of knowledge. In *Advances in Cryptology - CRYPTO '92, 12th Annual International Cryptology Conference, Santa Barbara, California, USA, August 16-20, 1992, Proceedings*, pages 390–420, 1992.

[BG08] Boaz Barak and Oded Goldreich. Universal arguments and their applications. *SIAM J. Comput.*, 38(5):1661–1694, 2008.

[BGGL01] Boaz Barak, Oded Goldreich, Shafi Goldwasser, and Yehuda Lindell. Resettably-sound zero-knowledge and its applications. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 116–125, 2001.

[BKP18] Nir Bitansky, Yael Tauman Kalai, and Omer Paneth. Multi-collision resistance: a paradigm for keyless hash functions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 671–684, 2018.

[BL18] Nir Bitansky and Huijia Lin. One-message zero knowledge and non-malleable commitments. In *TCC (1)*, volume 11239 of *Lecture Notes in Computer Science*, pages 209–234. Springer, 2018.

[Blu86] Manuel Blum. How to prove a theorem so no one else can claim it. *Proc of the International Congress of Mathematicians*, pages 1444–1451, 1986.

[CKLR11] Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz. Memory delegation. In *Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings*, pages 151–168, 2011.

[CLP13a] Ran Canetti, Huijia Lin, and Omer Paneth. Public-coin concurrent zero-knowledge in the global hash model. In *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings*, pages 80–99, 2013.

[CLP13b] Kai-Min Chung, Huijia Lin, and Rafael Pass. Constant-round concurrent zero knowledge from p-certificates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 50–59, 2013.

[COSV16] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. Concurrent non-malleable commitments (and more) in 3 rounds. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, pages 270–299, 2016.

[COSV17a] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. Delayed-input non-malleable zero knowledge and multi-party coin tossing in four rounds. In *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part I*, pages 711–742, 2017.

[COSV17b] Michele Ciampi, Rafail Ostrovsky, Luisa Siniscalchi, and Ivan Visconti. Four-round concurrent non-malleable commitments from one-way functions. In *Advances in Cryptology*

- CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part II, pages 127–157, 2017.

[DDN00] Danny Dolev, Cynthia Dwork, and Moni Naor. Nonmalleable cryptography. *SIAM J. Comput.*, 30(2):391–437, 2000.

[FLS99] Uriel Feige, Dror Lapidot, and Adi Shamir. Multiple noninteractive zero knowledge proofs under general assumptions. *SIAM J. Comput.*, 29(1):1–28, 1999.

[FS90] U. Feige and A. Shamir. Witness indistinguishable and witness hiding protocols. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, STOC 1990, May 13-17, 1990, Baltimore, Maryland, USA*, pages 416–426, 1990.

[GHV10] Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan. i-hop homomorphic encryption and rerandomizable yao circuits. In *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, pages 155–172, 2010.

[GK90] Oded Goldreich and Hugo Krawczyk. On the composition of zero-knowledge proof systems. In *Automata, Languages and Programming, 17th International Colloquium, ICALP90, Warwick University, England, July 16-20, 1990, Proceedings*, pages 268–282, 1990.

[GK96] Oded Goldreich and Hugo Krawczyk. On the composition of zero-knowledge proof systems. *SIAM J. Comput.*, 25(1):169–192, 1996.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. In *SIAM. Journal on Computing*, pages 186–208, 1989.

[GO94] Oded Goldreich and Yair Oren. Definitions and properties of zero-knowledge proof systems. *Journal of Cryptology*, 7(1):1–32, 1994.

[Gol01] Oded Goldreich. *The Foundations of Cryptography - Volume 1, Basic Techniques.* Cambridge University Press, 2001.

[GPR16] Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 1128–1141. ACM, 2016.

[GRRV14] Vipul Goyal, Silas Richelson, Alon Rosen, and Margarita Vald. An algebraic approach to non-malleability. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 41–50. IEEE Computer Society, 2014.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[HK12] Shai Halevi and Yael Tauman Kalai. Smooth projective hashing and two-message oblivious transfer. *J. Cryptology*, 25(1):158–193, 2012.

[Khu17] Dakshita Khurana. Round optimal concurrent non-malleability from polynomial hardness. In *Theory of Cryptography - 15th International Conference, TCC 2017,*, pages 139–171, 2017.

[KKS18] Yael Tauman Kalai, Dakshita Khurana, and Amit Sahai. Statistical witness indistinguishability (and more) in two messages. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings*, pages 34–65, 2018.

[KNY18] Ilan Komargodski, Moni Naor, and Eylon Yogev. Collision resistant hashing for paranoids: Dealing with multiple collisions. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings*, pages 162–194, 2018.

[KS17] Dakshita Khurana and Amit Sahai. How to achieve non-malleability in one or two rounds. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 564–575, 2017.

[LP11] Huijia Lin and Rafael Pass. Constant-round non-malleable commitments from any one-way function. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 705–714, 2011.

[LPS17] Huijia Lin, Rafael Pass, and Pratik Soni. Two-round and non-interactive concurrent non-malleable commitments from time-lock puzzles. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 576–587, 2017.

[LPTV10] Huijia Lin, Rafael Pass, Wei-Lung Dustin Tseng, and Muthuramakrishnan Venkitasubramaniam. Concurrent non-malleable zero knowledge proofs. In *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, pages 429–446, 2010.

[Nao91] Moni Naor. Bit commitment using pseudorandomness. *J. Cryptology*, 4(2):151–158, 1991.

[NP01] Moni Naor and Benny Pinkas. Efficient oblivious transfer protocols. In *Proceedings of the Twelfth Annual Symposium on Discrete Algorithms, January 7-9, 2001, Washington, DC, USA.*, pages 448–457, 2001.

[Pas13] Rafael Pass. Unprovable security of perfect NIZK and non-interactive non-malleable commitments. In *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings*, pages 334–354, 2013.

[Pas16] Rafael Pass. Unprovable security of perfect NIZK and non-interactive non-malleable commitments. *Computational Complexity*, 25(3):607–666, 2016.

[PR05a] Rafael Pass and Alon Rosen. Concurrent non-malleable commitments. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 563–572, 2005.

[PR05b] Rafael Pass and Alon Rosen. New and improved constructions of non-malleable cryptographic protocols. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, STOC 2005,*, pages 533–542, 2005.

[PRT13] Rafael Pass, Alon Rosen, and Wei-Lung Dustin Tseng. Public-coin parallel zero-knowledge for NP. *J. Cryptology*, 26(1):1–10, 2013.

[PTW09] Rafael Pass, Wei-Lung Dustin Tseng, and Douglas Wikström. On the composition of public-coin zero-knowledge protocols. In *Advances in Cryptology - CRYPTO 2009, 29th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2009. Proceedings*, pages 160–176, 2009.