2014 IEEE Annual Symposium on Foundations of Computer Science

An Algebraic Approach to Non-Malleabiliby

Vipul Goyal
Microsoft Research
Bangalore, India
vipul@microsoft.com

Silas Richelson
UCLA Comp. Sci. Dept.
Los Angeles, CA, USA
sirichel@ucla.edu

Alon Rosen
Efi Arazi School of Comp. Sci.
IDC Herzliya, Israel
alon.rosen@idc.ac.il

Margarita Vald
Blavatnik School of Comp. Sci.
Tel Aviv University, Israel
margarita.vald@cs.tau.ac.il

***Abstract*—In their seminal work on non-malleable cryptography, Dolev, Dwork and Naor, showed how to construct a non-malleable commitment with logarithmically-many "rounds"/"slots", the idea being that any adversary may successfully maul in some slots but would fail in at least one. Since then new ideas have been introduced, ultimately resulting in constant-round protocols based on any one-way function. Yet, in spite of this remarkable progress, each of the known constructions of non-malleable commitments leaves something to be desired.**

In this paper we propose a new technique that allows us to construct a non-malleable protocol with only a single "slot", and to improve in at least one aspect over each of the previously proposed protocols. Two direct byproducts of our new ideas are a four round non-malleable commitment and a four round non-malleable zero-knowledge argument, the latter matching the round complexity of the best known zero-knowledge argument (without the non-malleability requirement). The protocols are based on the existence of one-way functions and admit very efficient instantiations via standard homomorphic commitments and sigma protocols.

Our analysis relies on algebraic reasoning, and makes use of error correcting codes in order to ensure that committers' tags differ in many coordinates. One way of viewing our construction is as a method for combining many atomic sub-protocols in a way that simultaneously amplifies soundness and non-malleability, thus requiring much weaker guarantees to begin with, and resulting in a protocol which is much trimmer in complexity compared to the existing ones.

***Keywords*-non-malleability, commitments, zero-knowledge**

I. INTRODUCTION

The notion of non-malleability is central in cryptographic protocol design. Its objective is to protect against a man-in-the-middle (MIM) attacker that has the power to intercept messages and transform them in order to harm the security in other instantiations of the protocol. Commitment is often used as the paragon example for non-malleable primitives because of its ability to almost "universally" secure higher-level protocols against MIM attacks.

Commitments allow one party, called the committer, to probabilistically map a message m into a string, $\mathrm{Com}(m;r)$, which can be then sent to another party, called the receiver. In the statistically binding variant, the string $\mathrm{Com}(m;r)$ should be *binding*, in that it cannot be later "opened" into a message $m' \neq m$. It should also be *hiding*, meaning that for any pair of messages, m, m', the distributions $\mathrm{Com}(m;r)$ and $\mathrm{Com}(m';r')$ are computationally indistinguishable.

A commitment scheme is said to be *non-malleable* if for every message m, no MIM adversary, intercepting a commitment $\mathrm{Com}(m;r)$ and modifying it at will, is able to efficiently generate a commitment $\mathrm{Com}(\tilde{m};\tilde{r})$ to a related message $\tilde{m}$. Interest in non-malleable commitments is motivated both by the central role that they play in securing protocols under composition (see for example [CLOS02], [LPV09]) and by the unfortunate reality that many widely used commitment schemes are actually highly malleable. Indeed, man-in-the-middle (MIM) attacks occur quite naturally when multiple concurrent executions of protocols are allowed, and can be quite devastating.

Beyond protocol composition, non-malleable commitments are known to be applicable in secure multi-party computation [KOS03], [Wee10], [Goy11], authentication [NSS06], as well as a host of other non-malleable primitives (e.g., coin flipping, zero-knowledge, etc.), and even into applications as diverse as position based cryptography [CGMO09].

A. Prior Work

Since their conceptualization by Dolev, Dwork and Naor [DDN91], non-malleable commitments have been studied extensively, and with increasing success in terms of characterizing their round-efficiency and the underlying assumptions required. By now, we know how to construct constant-round non-malleable commitments based on any one-way function, and moreover the constructions are fully black-box. While this might give the impression that non-malleable commitments are well understood, each of the currently known constructions leaves something to be desired.

The first construction, due to DDN is perhaps the simplest and most efficient, mainly because it can in principle be instantiated with highly efficient cryptographic "sub-protocols". This, however, comes at the cost of round-complexity that is logarithmic in the maximum overall number of possible committers. Subsequent works, due to Barak [Bar02], Pass [Pas04], and, Pass and Rosen [PR05] are constant-round, but rely on (highly inefficient) non-black

0272-5428/14 $31.00 © 2014 IEEE
DOI 10.1109/FOCS.2014.13

box techniques. Wee [Wee10] (relying on [PW10]) gives a constant-round black-box construction under the assumption that sub-exponentially hard one-way functions exist. This construction employs a generic (and costly) transformation that is designed to handle general "non-synchronizing" MIM adversaries.

Finally, recent works by Goyal [Goy11] and Lin and Pass [LP11] attain non-malleable commitment with constant round-complexity via the minimal assumption that polynomial-time hard to invert one-way functions exist. The Lin-Pass protocol makes highly non-black-box use of the underlying one-way function (though not of the adversary), along with a concept called signature chains; resulting in significant overhead. Most relevant to the current work is the work of Goyal [Goy11]. Goyal's protocol, using a later result of Goyal, Lee, Ostrovsky and Visconti [GLOV12], can be made fully black-box, with its only shortcomings being high-communication complexity and the use of the Wee transformation (or alternatively a similarly costly transformation due to Goyal [Goy11]) for handling non-synchronizing adversaries. To construct non-malleable commitments, our work follows the blueprint proposed by Goyal, and introduces new proof techniques to significantly trim down its complexity, making various parts of the protocol of Goyal [Goy11] unnecessary.

The current state of affairs is such that in spite of all the remarkable advances, the DDN construction and its analysis remain the simplest and arguably most appealing candidate for non-malleable commitments. This is both due to its black-boxness and because it does not require transformations for handling a non-synchronizing MIM (in fact, the protocol is purposefully designed to introduce asynchronicity in message scheduling, which can be then exploited in the analysis).

B. Our Results

In this work we introduce a new algebraic technique for obtaining non-malleability, resulting in a simple and elegant non-malleable commitment scheme. The scheme's analysis contains many fundamentally new ideas allowing us to overcome substantial obstacles without sacrificing efficiency. The protocol is constructed using any statistically binding commitment scheme as a building block, and hence requires the minimal assumption that one way functions exist.

Theorem 1. *Assume the existence of a one-way functions. Then there is a 4-round non-malleable commitment scheme.*

Our protocol is appealing as in addition to requiring only the minimal assumption that one-way functions exist, it is much simpler and more efficient than all previous schemes. A direct consequence of our protocol is a 4-round non-malleable zero-knowledge argument based only on a OWF, demonstrating that for zero-knowledge, non-malleability does not necessarily come at the cost of extra rounds of interaction or complexity assumptions.

Theorem 2. *Suppose the existence of injective one-way functions. Then there is a 4-round black-box non-malleable zero-knowledge argument for every language in NP.*

Beyond the above virtues, we believe that our new techniques are actually the most significant contributions of this work. In addition to our use of algebra, we make novel combinatorial use of error correcting codes in order to ensure that different committers' tags differ in many coordinates. Whereas prior work relied on "worst-case" analysis of differences in committers' tags, ours follows from an "average-case" claim.

One way of viewing our construction is as a method for combining n atomic sub-protocols in a way that simultaneously amplifies their soundness and non-malleability properties, thus requiring much weaker soundness and non-malleability to begin with. We hope that this paradigm will become the norm for future work on in the area as, despite requiring more careful and strenuous analysis, it leads to pleasantly lightweight protocols. For example, this technique alone allows for an immediate linear reduction in communication complexity compared with its nearest relative, Goyal's protocol.

Another payoff of the algebraic techniques we employ is that our protocol only has one "slot". Nearly all of the non-malleable commitment schemes in the literature use multiple slots of interaction as a way to set up imbalances between the two different protocol instantiations that the MIM is involved in. The well known "two slot trick" of [Pas04], [PR05], [Goy11], for example, is a way to turn an arbitrary asymmetry between the instantiations into two: one which is heavy on the right and one on the left. The inability of the MIM to align the imbalances is crucial to the proof of non-malleability. Running the two slots in parallel introduces several technical problems, most notably "if the two imbalances are side by side, won't they just cancel each other out?" Our analysis uses a computational version of the "linear independence of polynomial evaluation" mantra in order to argue that the MIM cannot combine the two imbalances and must deal with each one separately.

We stress that the use of algebra and error correcting codes does not yield such reward for free: the analysis required becomes substantially more difficult. In the next section we describe and briefly discuss our new protocol and extractor. We then outline our techniques, keeping it informal but pointing out several of the challenges faced and new ideas required to overcome them.

C. The New Protocol

Suppose that committer C wishes to commit to message m, and let $t_1, \ldots, t_n \in \mathbb{Z}$ be a sequence of tags that uniquely correspond to C's identity (see Section II-A for a discussion of the tags). Let Com be a statistically binding commitment scheme, and suppose that $m \in \mathbb{F}_q$ where $q > \max_i 2^{t_i}$. The protocol proceeds as follows:

1) C chooses random $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{F}_q^n$ and sends $\mathrm{Com}(m)$ and $\{\mathrm{Com}(r_i)\}_{i=1}^n$ to R;
2) R a query vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \overset{\mathrm{R}}{\leftarrow} [2^{t_i}] \subset \mathbb{F}_q$;
3) C sends response $\mathbf{a} = (a_1, \ldots, a_n)$, $a_i = r_i \alpha_i + m$;
4) C proves in ZK that the values $\mathbf{a}$ (from step 3) are consistent with m and $\mathbf{r}$ (from step 1).

The statistical binding property of the protocol follows directly from the binding of Com. The hiding property follows from the hiding of Com, the zero-knowledge property of the protocol used in step 4, and from the fact that for every i the receiver R observes only a single pair of the form (α_i, a_i), where $a_i = r_i \alpha_i + m$.

Note the role of C's tags in the protocol: t_i determines the size of the $i-$th coordinate's challenge space. Historically, non-malleable commitment schemes have used the tags as a way for the committer to encode its identity into the protocol as a mechanism to prevent M (whose tag is different from C's tag) from mauling. In our protocol the tags play the same role, albeit rather passively. For example, though the size of the $i-$th challenge space depends on t_i, the size of the total challenge space depends only on the sum $\sum_{i=1}^n t_i$ of the tags. In particular, our scheme leaves open the possibility that the left and right challenge spaces might have the same size (in fact this will be ensured by our choice of tags). This raises a red flag, as previous works go to great lengths to set up imbalances between the left and right challenge spaces in order to force M to "give more information than it gets". Nevertheless, we are able to prove that any mauling attack will fail.

At a very high level, our protocol can be seen as an algebraic abstraction of Goyal's protocol. However, the fundamental difference we should emphasize from [Goy11] is that he crucially relies on the challenge space in the left interaction being much smaller than the challenge space in the right. For us, the challenge spaces in the two interactions are exactly the same size and so the techniques of [Goy11] do not apply to our setting−at least at first. Our protocol does have small imbalances between the challenge spaces of individual coordinates, which is what we will eventually use to prove non-malleability. However, proving that the coordinates are sufficiently independent so that these imbalances accrue to something usable is completely new to this work.

D. Proving Non-Malleability

Consider a MIM adversary M that is playing the role of the receiver in a protocol using tags $t_1, \ldots, t_n$ while playing the role of the committer in a protocol using tags $\tilde{t}_1, \ldots, \tilde{t}_n$. We refer to the former as the "left" interaction and to the latter as the "right" interaction. We let m and $\tilde{m}$ denote the messages committed to in the left and right interactions respectively. One nice feature of our protocol is that it is automatically secure against a non-synchronizing adversary, simply because there are so few rounds, there is no way for the MIM to benefit by changing the message order: any scheduling but the synchronous one can be dealt with trivially. So the only scheduling our proof actually needs to handle is a synchronizing one.

Our proof of non-malleability involves demonstrating the existence of an extractor, E, who is able to rewind M and extract $\tilde{m}$ without needing to rewind C in the left instantiation. Our extractor is modeled after Goyal's extractor which: (1) rewinds M to where $\tilde{\boldsymbol{\alpha}}$ was sent and asks a new query $\tilde{\boldsymbol{\beta}}$ instead, and (2) responds to M's left query randomly (it cannot do better without rewinding C as it does not know m), hoping that M answers correctly on the right.

In Goyal's protocol there is no way for E to know whether M answered correctly or not, and so it must have a verification message after the query response phase so E can compare M's answer with the main thread to verify correctness. We sidestep this necessity in the following way. We rewind to the beginning of step 2 twice and ask two new query vectors $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$, we answer randomly on the left obtaining $\big\{(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}}), (\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}), (\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{c}})\big\}$, where $(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{a}})$ is from the main thread. Comparing both $(\tilde{\beta}_i, b_i)$ and $(\tilde{\gamma}_i, c_i)$ with $(\tilde{\alpha}_i, a_i)$ will result in candidate values $\tilde{m}_i$ and $\tilde{m}_i'$, but with no verification message it is not clear how E should verify which one (if either) is correct. We accomplish this with the following "collinearity test". If $\tilde{m}_i = \tilde{m}_i'$ then E checks whether the points $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_i, \tilde{b}_i), (\tilde{\gamma}_i, \tilde{c}_i)\big\}$ are collinear. If so, E deems that $\tilde{m}_i$ was the correct value. This requires proving that M cannot answer "incorrectly but collinearly".

Tags in Error Corrected Form.: Just as in many of the existing NMC schemes, our protocol consists of n "atomic subprotocols", one for each tag. Previous schemes use the so called "DDN trick" [DDN91] in order to turn C's $k-$bit identity into a list of $n\ (= k)$ tags $t_1, \ldots, t_n$, satisfying the properties: (1) each t_i is of length $\log n + 1$; and (2) if $\{t_i\}_i$ and $\{\tilde{t}_j\}_j$ are the tags resulting from two distinct identities then there exists some i such that t_i is completely distinct from $\{\tilde{t}_j\}_j$, meaning that $t_i \neq \tilde{t}_j$ for all j.

Previous schemes' security proofs require the extractor to be able to use any completely distinct left subprotocol (i.e., one whose tag is completely distinct from $\{\tilde{t}_j\}_j$) to extract

M's commitment $\tilde{m}$ with high probability. This ensures that extraction is possible even in the worst case when there is a single such subprotocol. It also introduces a good deal of redundancy into the protocol.

While one would generally expect most pairs of distinct identities to result in pairs of tags such that property (2) holds for many i, all the DDN trick can guarantee in the worst case is that it holds for a single i (since M is allowed to choose his identity adversarily, this worst case situation might very well be realized). If however, one first applies an error correcting code to C's identity obtaining, say, a codeword in $\mathbb{F}^n$ for suitably chosen finite field $\mathbb{F}$ with $|\mathbb{F}| = \mathrm{poly}(n)$, then applying the DDN trick to this codeword would yield tags such that (1) t_i is of length $\mathcal{O}(\log n)$; and (2) t_i is completely distinct from $\{\tilde{t}_j\}_j$ for a constant fraction of the $i \in \{1, \ldots, n\}$.

Our "completely distinct on average" property requires only that extraction is possible from a completely distinct left subprotocol with constant probability, since there now are guaranteed to be many extraction opportunities. This allows us to remove much of the artificial redundancy resulting in an incredibly trim protocol.

Non-malleability against a copying M*:* To get a sense of why we might expect our scheme to be non-malleable, let us examine the situation against an M who attempts to maul C's commitment by simply copying its messages from the left interaction to the right. Let m be the message committed to on the left and let $\{t_i\}_{i=1}^n$ and $\{\tilde{t}_i\}_{i=1}^n$ be the corresponding tags.

After the first message, M will have copied C's commitments over to the right interaction, successfully committing to the coefficients of the linear polynomials $\tilde{f}_i(x) = r_i x + m$, $i = 1, \ldots, n$. The hiding of Com ensures it does not know the polynomials themselves, and so when it receives the right query vector $\tilde{\boldsymbol{\alpha}}$, its only hope of coming up with the correct valuations $\tilde{f}_i(\tilde{\alpha}_i)$ is to copy R's challenge to the left interaction and copy C's response back. However, it is unlikely that this will be possible. Indeed, M can only copy $\tilde{\alpha}_i$ over to the left when $\tilde{\alpha}_i \in \left[2^{t_i}\right]$. If $\tilde{t}_i > t_i$ then the $i-$th challenge space on the right is at least twice as big as the $i-$th challenge space on the left, which means that the probability $\tilde{\alpha}_i$ can be copied is at most $1/2$. We will use a code which ensures that $\tilde{t}_i > t_i$ for a constant fraction of the i, making the probability that M can copy every coordinate of R's query vector $\tilde{\boldsymbol{\alpha}}$ negligible. So M will not be able to successfully answer R's query and complete the proof when performing the "copying" attack.

Non-malleability against general M*:* Establishing security against a general man-in-the-middle adversary is significantly more challenging, and this is where the bulk of the new ideas are required. Our proof of non-malleability will require us to delve into the full range of possibilities for M's behavior. In each case, we will show that one of three things happen:

1) M does not correctly answer its queries with good enough probability;
2) E succeeds in extracting $\tilde{m}$ with sufficient probability;
3) an M with such behavior can be used to break the hiding of Com.

The core of our result can be seen as a reduction from a PPT M who correctly answers its queries with non-negligible probability and yet causes E to fail, to a machine $\mathcal{A}$ who breaks the hiding of Com. The following is a very high level outline of our proof.

We define USEFUL to be the set of transcripts which do not lead to situation 1 above; that is, transcripts for which M has a good chance of completing the protocol given the prefix. This is important in order for E to have any chance of successfully extracting $\tilde{m}$. Indeed, if M just aborts in every rewind, E will have no chance. From this standpoint, USEFUL is the set of transcripts which give E "something to work with." We prove that most transcripts are in USEFUL in Claim 3.

We then define EXT, the set of "extractable" transcripts, on which E will succeed with high probability. These are the transcripts which lead to situation 2. Intuitively, EXT is the set of transcripts such that M has good probability of correctly answering a query in a rewind despite the fact that E provides random answers to M's queries. We prove that indeed, if a transcript is in EXT then E succeeds in extracting $\tilde{m}$.

Finally, we define TRB, the set of "troublesome" transcripts which are both useful and not extractable. Transcripts in TRB are problematic as on the one hand, usefulness ensures that the prefix is such that if M receives correct responses to its queries on the left, it gives correct responses to the queries on the right. At the same time however, transcripts in TRB are not extractable and so the prefix is also such that if M receives random responses to its queries on the left it answers the right queries incorrectly. Certainly, the hiding of Com ensures that M cannot *know* whether it receives correct or random responses to its queries on the left. So this difference in behavior suggests that we may be able to use M to violate the hiding of Com, leading to situation 3 above.

Our main claim in this part of our proof is Claim 8, which says that if the left challenge $\boldsymbol{\alpha}$ has a superpolynomial number of preimage right challenges $\tilde{\boldsymbol{\alpha}}$ then either E succeeds in extracting $\tilde{m}$, or M can be used to break hiding. Such a claim has been at core of the analysis of some previous NMC schemes. In fact, as many previous schemes (such as [Goy11], for example) use multiple slots in order to *ensure* that some slot has a right challenge space that is much bigger than the left, such a claim often encompasses nearly the entire analysis. In our case, we have some work still left as there is only a single slot and the right and left challenge spaces have the same size. Nevertheless, we are able to prove, using a series of combinatorial arguments, that any mauling attack will wind up with M's left query having exponentially many preimage right queries.

To see these techniques in action, define the set $S = \{i \in [n] : \tilde{t}_i \le t_i\}$, and consider an M who simply copies the right challenges $\tilde{\alpha}_i$ for $i \in S$ over to the left but who makes sure to produce a legal query in the coordinates not in S on the left. As $\left[2^{\tilde{t}_i}\right] \subset \left[2^{t_i}\right]$ for all $i \in S$, copying $\tilde{\alpha}_i$ when $i \in S$ is fine. If we think of M as a map sending right challenge $\tilde{\boldsymbol{\alpha}}$ to left challenge $\boldsymbol{\alpha}$, then for any $\tilde{\boldsymbol{\alpha}}_S = (\tilde{\alpha}_i)_{i \in S}$, M sends $\tilde{\boldsymbol{\alpha}}'$ such that $\tilde{\boldsymbol{\alpha}}'_S = \tilde{\boldsymbol{\alpha}}_S$ to $\boldsymbol{\alpha}'$ such that $\boldsymbol{\alpha}'_S = \tilde{\boldsymbol{\alpha}}_S$. In other words, M maps the set of right query vectors whose $S-$coordinates are fixed to $\tilde{\boldsymbol{\alpha}}_S$ to the set of left query vectors whose $S-$coordinates are also fixed to $\tilde{\boldsymbol{\alpha}}_S$. However, the sizes of these subsets of right and left challenges are

$$\prod_{i \notin S} 2^{\tilde{t}_i} \text{ and } \prod_{i \notin S} 2^{t_i},$$

respectively, and $\prod_{i \notin S} 2^{\tilde{t}_i} = 2^{\Omega(n)} \prod_{i \notin S} 2^{t_i}$ (we are using that our tags are in error-corrected form, which ensures that $\left|[n] \setminus S\right| = \Omega(n)$). So we see that M, when restricted to the right challenges with $S-$coordinates fixed to $\tilde{\boldsymbol{\alpha}}_S$, is exponentially many to one on average, and so $\boldsymbol{\alpha}$ has exponentially many preimages with high probability.

4−Round Non-Malleability: The protocol described above is explained sequentially, and as written, consists of 8 rounds: two for Naor's commitment, two for the query/response phase, and four for the ZK argument. However, it can be parallelized down to four rounds using the Feige-Shamir four round ZK argument system [FS90]. This requires running the entire ZK argument in parallel with the commit, query and response messages. We make use of the fact that the statement to be proven can be chosen during the last round of the protocol, and that Feige-Shamir is actually an argument of knowledge, both of which have been used often in the literature. Armed with a 4-round NMC scheme, 4-round ZK is obtained essentially by running a 4-round ZK argument protocol (we again use Feige-Shamir) in parallel with a non-malleable commitment to the witness w. We point out, however, that by using Feige-Shamir we are assuming the existence of *injective* one-way functions.

Many-Many Non-Malleability: Many-many or concurrent non-malleability considers a setting where the MIM can run polynomially many protocols on the left and right (interleaved arbitrarily). It can be demonstrated to hold for our protocol using known techniques. First, one can show that our protocol is one-many non-malleable following [Goy11]. The key point is that the extractor we construct during our proof of non-malleability is able to extract $\tilde{m}$ from the right interaction with high probability, without rewinding the left execution. Therefore, by the union bound, our extractor will succeed in extracting from all of the right interactions with high probability. Next, we use the transformation of [LPV08], that one-many non-malleability implies many-many non-malleability. Their proof uses a hybrid argument to say that one-many non-malleability ensures that non-malleability is retained when adding polynomially many left executions, one by one.

II. Preliminaries

Let λ be the security parameter. For lack of space, we defer the definition of commitments and non-malleable commitments to the full version.

A. Tags in Error Corrected Form

Let $id \in \{0,1\}^k$ be C's identity and let $\mathbf{y} \in \mathbb{F}^{n/2}$ be the image of id under an error correcting code with constant distance, for a suitable finite field $\mathbb{F}$. Constant distance implies that if $id, \tilde{id} \in \{0,1\}^k$ are distinct identities then $\mathbf{y}$ and $\tilde{\mathbf{y}}$ differ on a constant fraction of their coordinates. Now, set

$$t_i = \begin{cases} 2i|\mathbb{F}| + y_i, & i \le n/2 \\ (2n+1)|\mathbb{F}| - t_{n-i}, & i > n/2 \end{cases}$$

Note that $2i|\mathbb{F}| \le t_i < (2i+1)|\mathbb{F}|$ for all i. The following is a list of useful properties that the tags satisfy. Let $\{t_i\}_i$ and $\{\tilde{t}_i\}_i$ be the tags resulting from distinct identities $id \ne \tilde{id}$.

1) $t_1 < t_2 < \cdots < t_n$;
2) $t_1 = \omega(\log \lambda)$ and $t_{i+1} - t_i = \omega(\log \lambda)$ for all $i \in [n]$; moreover $t_{i+1} - \tilde{t}_i = \omega(\log \lambda)$.
3) if $i \ne j$ then $t_i \ne \tilde{t}_j$; moreover $t_i < \tilde{t}_i$ holds for a constant fraction of $i \in [n]$ (as does $t_i > \tilde{t}_i$).

Properties 1 and 2 follow immediately as long as $|\mathbb{F}| = \omega(\log \lambda)$. Property 3 follows from 1) the distance of the error correcting code as $t_i = \tilde{t}_i$ iff $y_i = \tilde{y}_i$ which must not be the case for a constant fraction of the $i \in [n]$; along with 2) if $t_i \ne \tilde{t}_i$ then either $t_i < \tilde{t}_i$ or else $t_{n-i} < \tilde{t}_{n-i}$. This is reminiscent of the two slot trick of [Pas04], [PR05].

III. The Protocol

In this section, we describe our protocol given tags $t_1, \ldots, t_n$ in error corrected form as described in Section II-A. We use Naor's two round, statistically binding bit commitment scheme [Nao91] as a building block. We use boldface to denote vectors; in particular a challenge vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and a response vector $\mathbf{a} = (a_1, \ldots, a_n)$. We write $\mathbf{Com}$ for the entire first commitment message, so $\mathbf{Com} = \big(\mathrm{Com}(m), \mathrm{Com}(r_1), \ldots, \mathrm{Com}(r_n)\big)$. Our non-malleable commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle$ between a committer C trying to commit to m and a receiver R appears in Figure 1. The decommitment phase is done by having the committer C send m and the randomness it used during the protocol.

Public Parameters: Tags $t_1, \ldots, t_n$; prime $q > 2^{t_i}\ \forall i$.

Commiter's Private Input: Message $m \in \mathbb{F}_q$.

Commit Phase:

0) $\mathrm{R} \to \mathrm{C}$ **Initialization message:** Send the first message σ of the Naor commitment scheme.
1) $\mathrm{C} \to \mathrm{R}$ **Commit message:** Sample random $r_1, \ldots, r_n \in \mathbb{F}_q$ and $s, s_1, \ldots, s_n$.
   - Define linear functions $f_1, \ldots, f_n$ by $f_i(x) = r_i x + m$.
   - Send commitments $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \ldots, \mathrm{Com}_\sigma(r_n; s_n)\big)$.
2) $\mathrm{R} \to \mathrm{C}$ **Query:**
   - Send random challenge vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \in [2^{t_i}] \subset \mathbb{F}_q$.
3) $\mathrm{C} \to \mathrm{R}$ **Response:**
   - Send $\mathbf{a} = (a_1, \ldots, a_n)$, $a_i = f_i(\alpha_i)$.
4) $\mathrm{C} \longleftrightarrow \mathrm{R}$ **Consistency proof:** Parties engage in a zero-knowledge argument protocol where C proves to R that $\exists\ \big((m, s), (r_1, s_1), \ldots, (r_n, s_n)\big)$ such that:
   - $\mathbf{Com} = \big(\mathrm{Com}_\sigma(m; s), \mathrm{Com}_\sigma(r_1; s_1), \ldots, \mathrm{Com}_\sigma(r_n; s_n)\big)$; and
   - $a_i = r_i \alpha_i + m\ \forall\ i = 1, \ldots, n$.

Figure 1: The non-malleable commitment scheme $\langle \mathrm{C}, \mathrm{R} \rangle$.

Proposition 1. *$\langle \mathrm{C}, \mathrm{R} \rangle$ is a statistically binding commitment scheme.*

Theorem 3 (Main theorem)**.** *$\langle \mathrm{C}, \mathrm{R} \rangle$ is non-malleable.*

IV. Proof Sketch of Non-Malleability

In this section we prove Theorem 3. Just as with previous schemes, proving non-malleability amounts to constructing an extractor E who, given M's view after interacting with C on the left and R on the right, is able to extract M's commitment $\tilde{m}$ on the right without rewinding C on the left. The idea is that if E can extract $\tilde{m}$ without rewinding on the left then $\tilde{m}$ cannot depend in a meaningful way on m (the commitment on the left), as this would violate hiding. Our extractor is shown in Figure 2. The following theorem is sufficient for Theorem 3 (since it ensures that M breaks non-malleability AND E extracts with non-negligible probability).

Theorem 4. *Suppose* M *breaks the non-malleability of* $\langle \mathrm{C}, \mathrm{R} \rangle$ *with probability at least* $2p$ *for non-negligible* $p = p(\lambda)$. *We have* $\Pr_{\mathbb{T} \in \mathsf{ACC}}(\mathrm{E}(\mathbb{T}) \neq \tilde{m}) \leq p$.

Notation: We let $\mathbb{T}$ denote the transcript of interaction that E gets as input. Specifically,

$$\mathbb{T} = \big(\mathbf{Com}, \widetilde{\mathbf{Com}}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}, \tilde{\mathbf{a}}, \pi, \tilde{\pi}\big),$$

but as E is not interested in the proofs $(\pi, \tilde{\pi})$, and M is deterministic (and so $\widetilde{\mathbf{Com}}$, $\boldsymbol{\alpha}$, $\tilde{\mathbf{a}}$ are uniquely determined by $\mathbf{Com}$, $\tilde{\boldsymbol{\alpha}}$, and $\mathbf{a}$) we will often just write $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big)$.

Definition 1 (**Accepting Transcript**)**.** *We say that* $\mathbb{T} \in \mathsf{ACC}$ *if both* π *and* $\tilde{\pi}$ *are accepting proofs.*

The extractor E gets $\mathbb{T} \in \mathsf{ACC}$ as input so the probabilities which arise in our analysis often are conditioned on the event $\mathbb{T} \in \mathsf{ACC}$. We denote this with the convenient shorthand $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\cdots\big)$ instead of $\Pr_{\mathbb{T}}\big(\cdots \big| \mathbb{T} \in \mathsf{ACC}\big)$. For fixed $\mathbf{Com}$, M can be thought of as a deterministic map, mapping right query vectors to left ones. We write $\boldsymbol{\alpha} = \mathrm{M}(\tilde{\boldsymbol{\alpha}})$ to be consistent with this point of view. We assume that the transcript E gets as input is consistent with exactly one right commitment $\tilde{m}$. As $\langle \mathrm{C}, \mathrm{R} \rangle$ is statistically binding, this happens with overwhelming probability.

A. *The Extractor* E

The high level description of our extractor (described formally in Figure 2) is quite simple. Intuitively, our protocol begins by C committing to n, threshold 2, Shamir secret sharings [Sha79] of m; R then asks for one random share from each sharing, which C gives. All E does is rewind M to the beginning of the right session's query phase ask for a new random share. Since E gets one share as part of its input, this will allow E to reconstruct $\tilde{m}$.

The problem with this approach is that E does not know the value C has committed to on the left and so it does not know how to answer M's query on the left correctly. The best E can do is give a random response on the left and hope that M will give a correct response on the right anyway. On the one hand, the hiding of Com dictates that M cannot distinguish a correct response from a random one. On the other hand, M doesn't actually need to know whether the response on the left is correct or not in order to perform a successful mauling attack. Imagine, for example, the MIM who mauls R's challenge to the left execution and mauls C's response back. Such an M will prevent E from extracting $\tilde{m}$ because M only correctly answers E's query if given a correct response to its own left query, which E cannot give. Of course we will prove that no M with such behavior can exist, but this proof is highly non-trivial.

Another question which our extractor raises is "how can E tell a correct response from an incorrect one?" As we have described it, the hiding of Com ensures that it cannot. However, a small modification to the E described above fixes this. Instead of asking for one new share, E rewinds twice to the beginning of the right query phase and asks for two different new shares.

The key observation is that if M answers both queries correctly then the three shares it holds (the two it received plus the one it got as input) are collinear, whereas if M answers at least one incorrectly they are overwhelmingly likely to NOT be collinear. This is the first appearance of a tangeable payoff of the algebraicity of our protocol. For example, the protocol of [Goy11] (which is similar to ours, but strictly combinatorial in nature) does not have this algebraic verification technique at its disposal and must introduce use extra rounds into the protocol to ensure its extractor can reconstruct $\tilde{m}$.

Claim 1. *The probability that* E *answers a pair of queries* $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$ *"incorrectly but collinearly" is negligible.*

This claim ensures that if E fails, it is because M never answered a pair of queries correctly on the right.

Tags: Let $\{t_i\}_i$ and $\{\tilde{t}_i\}_i$ be in error corrected form.

Input: $\mathbb{T} = \big(\mathbf{Com}, \tilde{\boldsymbol{\alpha}}, \mathbf{a}\big) \in \mathsf{ACC}$, and a large value $N = \text{poly}(\lambda)$. E is given oracle access to M.

Extraction procedure: For $j \in [N]$:

1) Rewind M to the beginning of step 2 of the protocol:
   - generate a random right challenge vector $\tilde{\boldsymbol{\beta}}_j$.
   - Feed M with $\tilde{\boldsymbol{\beta}}_j$ and receive challenge $\boldsymbol{\beta}_j$.
2) Feed $\mathbf{b}_j = (b_{1,j}, \ldots, b_{n,j})$ to M where $b_{i,j}$ is random unless $\beta_{i,j} = \alpha_{i,j}$ in which case $b_{i,j} = a_i$. Receive $\tilde{\mathbf{b}}_j$.
3) For each $i \in [n]$ use $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j})\big\}$ to interpolate a line and recover candidate $\tilde{m}_{i,j}$.
4) Repeat steps 1-3. Let $\tilde{\boldsymbol{\gamma}}_j$ be new right challenge vector, $\tilde{\mathbf{c}}_j$ the response and $(\tilde{m}'_{1,j}, \ldots, \tilde{m}'_{n,j})$ the recovered candidates.
5) If for some $i \in [n]$, $\tilde{m}_{i,j} = \tilde{m}'_{i,j}$ and $\big\{(\tilde{\alpha}_i, \tilde{a}_i), (\tilde{\beta}_{i,j}, \tilde{b}_{i,j}), (\tilde{\gamma}_{i,j}, \tilde{c}_{i,j})\big\}$ are collinear output $\tilde{m}_{i,j}$ and halt.

Output: Output **FAIL**.

Figure 2: The Extractor E.

B. Extractable, Useful and Troublesome Transcripts

Our extractor E is parametrized by a large polynomial $N = N(\lambda)$, which is the number of times E rewinds. Specifically, setting $N = \omega(\lambda n^{10} p^{-18})$ suffices for our proof. We remark that there is no reason to suspect that N must be such a large polynomial; it arises from our analysis, which is not concerned with minimizing N.

Definition 2 (Extractable Transcripts). *Fix non-negligible* $\varepsilon^* = \big(\lambda/N\big)^{1/2}$. *We define* EXT *as the* $\mathbb{T} \in \mathsf{ACC}$ *such that for some* $i \in [n]$, *the probability that* M *answers* $\tilde{\beta}_i$ *correctly on the right given that his left queries are answered by* E *is at least* ε^*.

Claim 2. $\Pr_{\mathbb{T}}\big(\mathrm{E}(\mathbb{T}) = \mathbf{FAIL} \big| \mathbb{T} \in \mathsf{EXT}\big) = \mathbf{negl}(\lambda)$, *where the probability is over* $\mathbb{T}$ *and the randomness of* E.

Definition 3 (Useful Transcripts). *Fix non-negligible* $\delta < \frac{1}{3}$ *and (temporarily) define*

$$W = \big\{\mathbf{Com} : \Pr_{\mathbb{T}}\big(\mathbb{T} \in \mathsf{ACC} \big| \mathbf{Com}\big) \leq \delta p^2\big\}.$$

Set $\mathsf{USEFUL} := \big\{\mathbb{T} \in \mathsf{ACC} : \mathbf{Com} \notin W\big\}$.

Claim 3. $\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \notin \mathsf{USEFUL}\big) \leq \delta p$.

Transcripts in EXT are those for which M is likely to correctly answer a right query even given incorrect responses to its own left queries. On the other hand, USEFUL can be thought of as the transcripts for which M answers the right queries correctly if given correct answers to its left queries. This leads us to the following definition.

Definition 4 (Troublesome Transcripts). *We define* $\mathsf{TRB} = \mathsf{USEFUL} \setminus \mathsf{EXT}$.

Transcripts in TRB are troublesome as essentially, they are transcripts for which M answers the right queries correctly iff his left queries are answered correctly. However, the hiding of Com ensures that M cannot *know* whether it receives correct or random responses to its queries on the left. So this difference in behavior suggests that we may be able to use M to break the hiding of Com. However, it is not so easy. Keep in mind, M does not have to know whether it is giving a correct or incorrect answer on the right in order to successfully maul. Indeed, almost all mauling attacks one could imagine have the property that M answers correctly on the right if and only if it gets correct answers on the left. The following lemma comprises the heart of our analysis.

Lemma 1. *If* Com *is computationally hiding then there exists a constant* $\delta' < \frac{1}{3}$ *such that*

$$\Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB}\big) \leq \delta' p.$$

Lemma 1 combined with Claims 1 through 3 give us

$$\begin{aligned} \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathrm{E}(\mathbb{T}) \neq \tilde{m}\big) \quad &\leq \quad \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \notin \mathsf{USEFUL}\big) \\ &+ \quad \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\mathbb{T} \in \mathsf{TRB}\big) \\ &+ \quad \Pr_{\mathbb{T}}\big(\mathrm{E}(\mathbb{T}) = \mathbf{FAIL} \big| \mathbb{T} \in \mathsf{EXT}\big) \\ &+ \quad \Pr_{\mathbb{T} \in \mathsf{ACC}}\big(\text{inc. but coll.}\big) \\ &\leq \quad \delta p + \delta' p + \mathbf{negl}(\lambda) < p, \end{aligned}$$

proving Theorem 4.

C. Proof Overview of Lemma 1

We complete our discussion of the proof of Theorem 3 with a high level overview of the proof of Lemma 1. We leave out the formal proofs of the necessary claims because of space. An interested reader should consult the full version. We will, however, in this discussion state the claims, numbered as in the full version to make the transition as easy as possible (this is why we skip over Claim 4).

We prove Lemma 1 by defining the notion of "query dependence", and then considering the possible different ways in which M's left queries $\boldsymbol{\alpha}$ can depend on right queries $\tilde{\boldsymbol{\alpha}}$. Intuitively, $\alpha_{i'}$ being dependent on $\tilde{\alpha}_i$ is the result of M performing a mauling attack. Suppose that M mauls $\mathrm{Com}(f_{i'})$ in order to obtain $\mathrm{Com}(\tilde{f}_i)$. Then M does not know $\tilde{f}_i$ and so cannot hope to answer $\tilde{\alpha}_i$ except by mauling C's answer to $\alpha_{i'}$. Therefore, if M is rewound to the beginning of step 2 and asked a different query vector $\tilde{\boldsymbol{\beta}}$ such that $\tilde{\beta}_i = \tilde{\alpha}_i$, M will have to ask $\boldsymbol{\beta}$ such that $\beta_{i'} = \alpha_{i'}$ if it wants to answer successfully. This is the idea of query dependence: if $\tilde{\alpha}_i$ is asked on the right, then $\alpha_{i'}$ must be asked on the left.

Recall that in the introduction we considered a copying MIM who attempts to maul C's commitment by simply copying and pasting messages between the left and right sessions. Such an attack is a very simple example of a mauling attack in which each α_i is dependent on $\tilde{\alpha}_i$. We saw this attack is foiled by the large number of left tags which differ from all right tags, preventing the right query $\tilde{\boldsymbol{\alpha}}$ from being a legal left query except with negligible probability. In fact, we prove in Claim 7 that all mauling attacks in which each α_i depends on $\tilde{\alpha}_i$ will fail whp.

This encourages us to investigate what else can happen. We arrive at three possibilities.

- UNBAL**:** There exist $i' > i$ such that $\alpha_{i'}$ depends on $\tilde{\alpha}_i$.
- $\mathsf{1-2}$**:** There exist (i_1, i_2, i') such that $\alpha_{i'}$ depends on both $\tilde{\alpha}_{i_1}$ and $\tilde{\alpha}_{i_2}$.
- IND**:** There exists i such that each $\alpha_{i'}$ does *not* depend on $\tilde{\alpha}_i$.

In the actual proof we formalize the above possibilities using precise conditional probability statements. We keep it informal here, however, in order to convey as much intuition as possible.

Note that if none of the above three events occur then α_i depends on $\tilde{\alpha}_i$ for all i which is what we hope happens. We complete the proof by showing that each of the three events cannot happen except with very small probability. However, this is easier said than done. Consider, for example, the mauling attack which results in $1-2$. Intuitively, if $\alpha_{i'}$ is dependent on both $\tilde{\alpha}_{i_1}$ and $\tilde{\alpha}_{i_2}$ then M is using C's response $f_{i'}(\alpha_{i'})$ on the left to produce both $\tilde{f}_{i_1}(\tilde{\alpha}_{i_1})$ and $\tilde{f}_{i_2}(\tilde{\alpha}_{i_2})$ on the right. On the one hand it is extremely unlikely that a single polynomial evaluation on the left contains enough information to allow M to correctly give two random evaluations on the right. On the other hand, this intuition alone isn't enough to say that $1-2$ can't occur as the argument is information theoretic in nature. Indeed, any statment one wishes to make about M's behavior in the query phase must have a computational proof as an unbounded M can query however it wants to and then simply break the hiding of the commitments in the first message to learn the $\tilde{f}_i$ and answer correctly.

The key claim which allows us to capitalize on our information theoretic intuition is Claim 8 which states that if the left query $\boldsymbol{\alpha}$ has a superpolynomial number of preimage right queries $\tilde{\boldsymbol{\alpha}}$ then either E succeeds in extracting $\tilde{m}$ or M can be used to break the hiding of $\langle \mathrm{C}, \mathrm{R} \rangle$. The intuition is that if there are superpolynomially many $\tilde{\boldsymbol{\alpha}}$ such that $\mathrm{M}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\alpha}$, the chances that M can use C's response by itself to answer $\tilde{\boldsymbol{\alpha}}$ are negligible. It follows that either M must be content to not answer most of the $\tilde{\boldsymbol{\alpha}}$ such that $\mathrm{M}(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{\alpha}$ (the probability of which can be bounded using a straightforward conditional probability argument) or M must know some "extra information" about the $\tilde{f}_i$ which allows him to provide a correct response to $\tilde{\boldsymbol{\alpha}}$. But this means that either M will use this extra information to correctly answer $\tilde{\boldsymbol{\alpha}}$ even when given a random answer to $\boldsymbol{\alpha}$ on the left (in which case E succeeds in extracting $\tilde{m}$), or M is choosing to utilize this extra information only when C answers correctly on the left. However, the hiding of the commitment in the first message ensures that M cannot *know* whether he receives correct responses on the left or not, and this difference in behavior will allow us to use M to break hiding.

Armed with Claim 8, we can now make definitive statements about UNBAL and $1-2$. For example, if UNBAL occurs then $\alpha_{i'}$ is dependent on $\tilde{\alpha}_i$ for some $i' > i$, and so if R asks a new right challenge with the same $i-$th query, M will fix $\alpha_{i'}$ on the left. However, as $i' > i$, $\alpha_{i'}$ is drawn from a much larger challenge space than $\tilde{\alpha}_i$, and so M is "wasting challenge space". Specifically, the residual right challenge space with the $i-$th query fixed to $\tilde{\alpha}_i$ is superpolynomially larger than the residual left challenge space with $\alpha_{i'}$ fixed, and so with high probability, we will find ourselves in a situation where the left query has superpolynomially many right query preimages. By Claim 8, this must not happen except with negligible probability. In this spirit, we prove Claims 5 through 7 which show that if either UNBAL or $1-2$ or "not (UNBAL or $1-2$ or IND)" occur, then the left query will have superpolynomially many right query preimages. The proofs of Claims 6 and 7 are more involved than that of Claim 5, but they are still purely combinatorial.

Finally, we prove in Claim 9 that IND cannot happen using another reduction to hiding. It uses the same framework as Claim 8 and has similar underlying intuition. Here the main point is that if IND occurs then there exists a right query $\tilde{\alpha}_i$ on which no $\alpha_{i'}$ on the left is dependent. Intuitively this means that M does not need any of the left challenges in order to correctly return $\tilde{f}_i(\tilde{\alpha}_i)$, implying that he knows some information about the polynomial $\tilde{f}_i$. As in the intuition for Claim 8 this means either that extraction is successful, or that M is breaking hiding.

D. Statements of Claims

Let $\sigma = \sigma(\lambda)$ be a non-negligible quantity defined for convenience (explicitly we set $\sigma = \varepsilon'(\delta')^2p^4/257n^3$, because of calculations in the full version).

Claim 5. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{UNBAL}\big)\geq\frac{\delta' p}{4}$, *then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{SUPER{-}POLY}\big)\geq\sigma.$$

Claim 6. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap 1{-}2\big)\geq\frac{\delta' p}{4}$, *then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{SUPER{-}POLY}\big)\geq\sigma.$$

Claim 7. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\setminus(\mathsf{UNBAL}\cup 1{-}2\cup\mathsf{IND})\big)\geq\frac{\delta' p}{4}$, *then*

$$\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{SUPER{-}POLY}\big)\geq\sigma.$$

Claim 8. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{SUPER{-}POLY}\big)\geq\sigma$ *then there exists a PPT algorithm $\mathcal{A}$ who breaks the hiding of $\langle\mathrm{C},\mathrm{R}\rangle$.*

Claim 9. *If* $\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\big)\geq\frac{\delta' p}{4}$ *then there exists a PPT algorithm $\mathcal{A}$ who breaks the hiding of $\langle\mathrm{C},\mathrm{R}\rangle$.*

Claims 5 through 9 combine to give that if Com is computationally hiding, then

$$\begin{aligned}\Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\big) &\leq \Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{UNBAL}\big)\\ &+ \Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap 1{-}2\big)\\ &+ \Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\cap\mathsf{IND}\big)\\ &+ \Pr_{\mathbb{T}\in\mathsf{ACC}}\big(\mathbb{T}\in\mathsf{TRB}\setminus(\mathsf{UNBAL}\cup\\ &\cup 1{-}2\cup\mathsf{IND})\big)\\ &\leq \frac{\delta' p}{4}+\frac{\delta' p}{4}+\frac{\delta' p}{4}+\frac{\delta' p}{4}=\delta' p,\end{aligned}$$

completing the proof of Lemma 1, Theorem 4 and Theorem 3.

V. Non-Malleability in 4-Rounds

In this section we show how to squeeze our non-malleable protocol $\langle\mathrm{C},\mathrm{R}\rangle$ into 4 rounds. In the new protocol, the zero-knowledge messages are lifted up and sent together with the commit, challenge and response messages. We use a version of the $4-$round zero-knowledge argument of knowledge protocol of Feige and Shamir [FS90] which can be constructed from a OWF. Such a protocol has been used before (see [**?**], for example). Recall briefly that in the Feige-Shamir protocol, V sets a trapdoor using a 3-round witness-hiding argument of knowledge, π_1 and then P uses a 3-round witness-indistinguishable argument of knowledge, π_2 to prove either the original statement or knowledge of V's trapdoor. Our 4-round commitment scheme $\langle\mathrm{C},\mathrm{R}\rangle_{\mathsf{OPT}}$ appears in Figure 3.

Public Parameters: Tags $t_1,\ldots,t_n$, and prime $q>2^{t_i}\ \forall i$.

Commiter's Private Input: Message $m\in\mathbb{F}_q$ to be committed to.

1) $\mathrm{R}\rightarrow\mathrm{C}$: Sample random $x_1\in L_1$ and $x'_1\in L'_1$ together with witnesses w_1 and w'_1, and send the first message of π_1 proving that $x_1\in L_1\bigvee x'_1\in L'_1$, along with the first message σ of Naor's commitment scheme.
2) $\mathrm{C}\rightarrow\mathrm{R}$: Send the challenge message of π_1 along with the first message of π_2 and the commitment message $\mathbf{Com}$ from Step 1 of $\langle\mathrm{C},\mathrm{R}\rangle$ The statement of π_2 will be determined in step 4.
3) $\mathrm{R}\rightarrow\mathrm{C}$: Send the last message of π_1 along with the second message of π_2 and the challenge vector $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)$ as done in Step 2 of $\langle\mathrm{C},\mathrm{R}\rangle$.
4) $\mathrm{C}\rightarrow\mathrm{R}$: Send the evaluation vector $\mathbf{a}$ where $a_i=r_i\alpha_i+m$ as in Step 3 of $\langle\mathrm{C},\mathrm{R}\rangle$ along with the last message of π_2 proving the statement:
   - EITHER: $\exists\ \big((m;s),(r_1;s_1),\ldots,(r_n;s_n)\big)$ such that $\mathbf{Com}$ and $\mathbf{a}$ are correct
   - OR: $x_1\in L_1\bigvee x'_1\in L'_1$.

Figure 3: : 4-round NMC scheme $\langle\mathrm{C},\mathrm{R}\rangle_{\mathsf{OPT}}$.

Proposition 2. *Assume the existence of OWFs. Then $\langle\mathrm{C},\mathrm{R}\rangle_{\mathsf{OPT}}$ is a $4-$round statistically binding, non-malleable commitment scheme.*

Using our new commitment scheme $\langle\mathrm{C},\mathrm{R}\rangle_{\mathsf{OPT}}$, we obtain a simple $4-$round non-malleable zero knowledge argument $\langle\mathrm{P},\mathrm{V}\rangle$ for any language $L\in\mathcal{NP}$. A detailed description of $\langle\mathrm{P},\mathrm{V}\rangle$ appears in the full version. The basic idea is to run a $4-$round ZK in parallel with $\langle\mathrm{C},\mathrm{R}\rangle_{\mathsf{OPT}}$ which P uses to commit non-malleably to his witness.

Proposition 3. *Assume the existence of OWFs. Then $\langle\mathrm{P},\mathrm{V}\rangle$ is a $4-$round non-malleable zero knowledge argument of knowledge for any $L\in\mathcal{NP}$.*

Acknowledgment

We would like to thank the anonymous FOCS referees for very helpful comments and suggestions which helped us improve our writeup significantly.

References

[Bar02] Boaz Barak. Constant-Round Coin-Tossing with a Man in the Middle or Realizing the Shared Random String Model. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '02, pages 345–355, 2002.

[CGMO09] Nishanth Chandran, Vipul Goyal, Ryan Moriarty, and Rafail Ostrovsky. Position based cryptography. In Shai Halevi, editor, *CRYPTO*, volume 5677 of *Lecture Notes in Computer Science*, pages 391–407. Springer, 2009.

[CLOS02] Ran Canetti, Yehuda Lindell, Rafail Ostrovsky, and Amit Sahai. Universally composable two-party and multi-party secure computation. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, STOC '02, pages 494–503, 2002.

[DDN91] Danny Dolev, Cynthia Dwork, and Moni Naor. Non-Malleable Cryptography (Extended Abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, STOC '91, pages 542–552, 1991.

[FS90] Uriel Feige and Adi Shamir. Witness indistinguishable and witness hiding protocols. In *STOC*, pages 416–426. ACM, 1990.

[GLOV12] Vipul Goyal, Chen-Kuei Lee, Rafail Ostrovsky, and Ivan Visconti. Constructing non-malleable commitments: A black-box approach. In *FOCS*, pages 51–60. IEEE Computer Society, 2012.

[Goy11] Vipul Goyal. Constant Round Non-malleable Protocols Using One-way Functions. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 695–704. ACM, 2011.

[KOS03] Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Round Efficiency of Multi-party Computation with a Dishonest Majority. In *Advances in Cryptology — EUROCRYPT '03*, volume 2656 of *Lecture Notes in Computer Science*, pages 578–595. Springer, 2003.

[LP11] Huijia Lin and Rafael Pass. Constant-round Non-malleable Commitments from Any One-way Function. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 705–714, 2011.

[LPV08] Huijia Lin, Rafael Pass, and Muthuramakrishnan Venkitasubramaniam. Concurrent Non-malleable Commitments from Any One-Way Function. In *Theory of Cryptography, 5th Theory of Cryptography Conference, TCC 2008*, pages 571–588, 2008.

[LPV09] Huijia Lin, Rafael Pass, and Muthuramakrishnan Venkitasubramaniam. A Unified Framework for Concurrent Security: Universal Composability from Stand-alone Non-malleability. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, STOC '09, pages 179–188, 2009.

[Nao91] Moni Naor. Bit Commitment Using Pseudorandomness. *J. Cryptology*, 4(2):151–158, 1991.

[NSS06] Moni Naor, Gil Segev, and Adam Smith. Tight bounds for unconditional authentication protocols in the manual channel and shared key models. In Cynthia Dwork, editor, *CRYPTO*, volume 4117 of *Lecture Notes in Computer Science*, pages 214–231. Springer, 2006.

[Pas04] Rafael Pass. Bounded-Concurrent Secure Multi-Party Computation with a Dishonest Majority. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, STOC '04, pages 232–241, 2004.

[PR05] 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 '05, pages 533–542, 2005.

[PW10] Rafael Pass and Hoeteck Wee. Constant-Round Non-malleable Commitments from Sub-exponential One-Way Functions. In *Advances in Cryptology — EUROCRYPT '10*, pages 638–655, 2010.

[Sha79] Adi Shamir. How to Share a Secret. *Commun. ACM*, 22(11):612–613, 1979.

[Wee10] Hoeteck Wee. Black-Box, Round-Efficient Secure Computation via Non-malleability Amplification. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2010.