Designs, Codes and Cryptography (2023) 91:3457–3525
https://doi.org/10.1007/s10623-023-01245-1

Zero-knowledge proofs for set membership: efficient, succinct, modular

Daniel Benarroch[1,6] · Matteo Campanelli[2,6] · Dario Fiore[3,6] · Kobi Gurkan[4,5,6] · Dimitris Kolonelos[2,3,6]

Received: 6 July 2022 / Revised: 1 March 2023 / Accepted: 3 May 2023 /
Published online: 1 July 2023

Abstract
We consider the problem of proving in zero knowledge that an element of a public set satisfies a given property without disclosing the element, i.e., for some u, "$u \in S$ and $P(u)$ holds". This problem arises in many applications (anonymous cryptocurrencies, credentials or whitelists) where, for privacy or anonymity reasons, it is crucial to hide certain data while ensuring properties of such data. We design new *modular* and *efficient* constructions for this problem through new *commit-and-prove zero-knowledge systems for set membership*, i.e. schemes proving $u \in S$ for a value u that is in a public commitment c_u. We also extend our results to support *non-membership proofs*, i.e. proving $u \notin S$. Being commit-and-prove, our solutions can act as plug-and-play modules in statements of the form "$u \in S$ and $P(u)$ holds" by combining our set (non-)membership systems with any other commit-and-prove scheme for $P(u)$. Also, they work with Pedersen commitments over prime order groups which makes them compatible with popular systems such as Bulletproofs or Groth16. We implemented our schemes as a software library, and tested experimentally their performance. Compared to previous work that achieves similar properties—the clever techniques combining zkSNARKs and Merkle Trees in Zcash—our solutions offer more flexibility, shorter public parameters and $3.7\times$–$30\times$ faster proving time for a set of size 2^{64}.

Keywords Public-key cryptography · Zero-knowledge proofs · Applications

Mathematics Subject Classification 94A60

This is one of several papers published in *Designs, Codes and Cryptography* comprising the "Special Issue: Mathematics of Zero Knowledge".

✉ Dimitris Kolonelos
dimitris.kolonelos@imdea.org

1 Inversed Technologies, Madrid, Spain

2 Protocol Labs, San Francisco, USA

3 IMDEA Software Institute, Madrid, Spain

4 Universidad Politécnica de Madrid, Madrid, Spain

5 Geometry Research, London, UK

6 cLabs, Berlin, Germany

1 Introduction

The problem of proving set membership—that a given element x belongs to some set S—arises in many applications, including governmental white-lists to prevent terrorism or money-laundering, voting and anonymous credentials, among others. More recently, this problem also appears at the heart of currency transfer and identity systems over blockchains. In this setting, parties can first publicly commit to sets of data (through the blockchain itself) and then, by proving set membership, can claim ownership of assets or existence of identity attributes, while ensuring privacy.

A naive approach to check if an element is in a set is to go through all its entries. The complexity of this approach, however, is unacceptable in many scenarios. This is especially true for blockchains, where most of the parties (the verifiers) should run quickly.

How to efficiently verify set membership then? Cryptographic *accumulators* [6] provide a nice solution to this problem. They allow a set of elements to be compressed into a short value (the accumulator) and to generate membership proofs that are short and fast to verify. As a security guarantee they require it should be computationally infeasible to generate a false membership proof.

As of today, we can divide constructions for accumulators into three main categories: Merkle Trees [55]; RSA-based [2, 11, 16, 50]; pairing-based [17, 32, 57, 78]. Approaches based on Merkle Trees[1] allow for short (i.e., $O(1)$) public parameters and accumulator values, whereas the witness for membership proofs is of size $\log(n)$, where n is the size of the set. In RSA-based constructions (which can be actually generalized to any group of unknown order [48], including class groups) both the accumulator and the witness are each a single element in a relatively large hidden-order group $\mathbb{G}$,[2] and thus of constant-size. Schemes that use pairings in elliptic curves such as [17, 57] offer small accumulators and small witnesses (which can each be a single element of a prime order bilinear group, e.g., 256 bits) but require large parameters (approximately $O(n)$) and a trusted setup.

In anonymous cryptocurrencies, e.g. Zerocash [5] (but also in other applications such as Anonymous Credentials [22] and whitelists), we also require *privacy*. That is, parties in the system would not want to disclose *which* element in the set is being used to prove membership. Phrased differently, one desires to prove that $u \in S$ without revealing u, or: the proof should be *zero-knowledge* [45] for u. As an example, in Zerocash users want to prove that a coin exists (i.e. belongs to the set of previously sent coins) without revealing which coin it is that they are spending.

In practice it is common that this privacy requirement goes beyond proving membership. In fact, these applications often require proving further properties about the accumulated elements, e.g., that for some element u in the set, property $P(u)$ holds. And this without leaking any more information about u other than what is entailed by P. In other words, we desire zero-knowledge for the statement $R^*(S, u) :=$ "$u \in S$ and $P(u)$".

One way to solve the problem, as done in Zerocash, is to directly apply general-purpose zero-knowledge proofs for R^*, e.g., [46, 61]. This approach, however, tends to be expensive and ad-hoc. One of the questions we aim to tackle is that of providing a more efficient proof systems for set membership relations, that can also be modular.

Specifically, as observed in [18], the design of practical proof systems can benefit from a more modular vision. A modular framework such as the one in [18] not only allows for

[1] We can include under this class currently known lattice-based accumulators such as [9, 60].

[2] The group $\mathbb{G}$ is typically $\mathbb{Z}_N^*$ where N is an RSA modulus. The size of an element in this group for a standard 128-bit security parameter is of 3072 bits.

separation of concerns, but also increases reusability and compatibility in a plug-and-play fashion: the same proof system is designed once and can be reused for the same sub-problem regardless of the context[3]; it can be replaced with a component for the same sub-problem at any time. Also, as [18] shows, this can have a positive impact on efficiency since designing a special-purpose proof system for a specific relation can lead to significant optimizations. Finally, this compositional approach can also be leveraged to build general-purpose proof systems.

In this work we focus on applying this modular vision to designing *succinct zero-knowledge proofs for set membership*. Following the abstract framework in [18] we investigate how to apply commit-and-prove techniques [20] to our setting. Our approach uses commitments for composability as follows. Consider an efficient zero-knowledge proof system Π for property $P(u)$. Let us also assume it is commit-and-prove, i.e. the verifier can test $P(u)$ by simply holding a commitment $c(u)$ to u. Such Π could be for example a commit-and-prove NIZK such as Bulletproofs [13] or a commit-and-prove zkSNARK such as LegoGroth16 from [18] that are able to operate on Pedersen commitments $c(\cdot)$ over elliptic curves. In order to obtain a proof gadget for set membership, all one needs to design is a commit-and-prove scheme for the relations "$u \in S$" where *both* u and S are committed: u through $c(u)$ and S through some other commitment for sets, such as an accumulator.

Our main contribution is to propose a formalization of this approach and new constructions of succinct zero-knowledge commit-and-prove systems for set membership. In addition, as we detail later, we also extend our results to capture proofs of *non-membership*, i.e., to show that $u \notin S$. For our constructions we focus on designing schemes where $c(u)$ is a Pedersen commitment in a prime order group $\mathbb{G}_q$. We focus on linking through Pedersen commitments as these can be (re)used in some of the best state-of-the-art zero-knowledge proof systems for general-purpose relations that offer for example the shortest proofs and verification time (see, e.g., [46] and its efficient commit-and-prove variant [18]), or transparent setup and logarithmic-size proofs [13].

Before describing our results in more detail, we review existing solutions and approaches to realize commit-and-prove zkSNARKs for set membership.

1.1 Existing approaches for proving set membership for pedersen commitments

The accumulator of Nguyen [57], by the simple fact of having a succinct pairing-based verification equation, can be combined with standard zero-knowledge proof techniques (e.g., Sigma protocols or the celebrated Groth–Sahai proofs [47]) to achieve a succinct system with reasonable proving and verification time. The main drawbacks of using [57], however, are the large public parameters (i.e. requiring as many prime group elements as the elements in the set) and a high cost for updating the accumulator to the set, in order to add or remove elements (essentially requiring to recompute the accumulator from scratch).

By using general-purpose zkSNARKs one can obtain a solution with constant-size proofs based on Merkle Trees: prove that there exists a valid path which connects a given leaf to the root; this requires proving correctness of about $\log n$ hash function computations (e.g., SHA256). This solution yields a constant-size proof and requires $\log n$-size public parameters if one uses preprocessing zkSNARKs such as [46, 61]. On the other hand, often when proving a relation such as $R^*(S, u) :=$ "$u \in S$ and $P(u)$" the bulk of the work stems

[3] For instance, one can plug a proof system for matrix product $C = A \cdot B$ in any larger context of computation involving matrix multiplication. This regardless of whether, say, we then hash C or if A, B are in turn the output of a different computation.

from the set membership proof. This is the case in Zcash or Filecoin[4] where the predicate $P(\cdot)$ is sufficiently small.

Finally, another solution that admits constant-size public parameters and proofs is the protocol of [16]. Specifically, Camenisch and Lysyanskaya showed how to prove in zero-knowledge that an element u committed in a Pedersen commitment over a prime order group $\mathbb{G}_q$ is a member of an RSA accumulator. In principle this solution would fit the criteria of the gadget we are looking for. Nonetheless, its concrete instantiations show a few limitations in terms of efficiency and flexibility. The main problem is that, for its security to hold, we need a prime order group (the commitment space) and the primes (the message space) to be quite large, for example[5] $q > 2^{519}$. But having such a large prime order group may be undesirable in practice for efficiency reasons. In fact the group $\mathbb{G}_q$ is the one that is used to instantiate more proof systems that need to interact and be linked with the Pedersen commitment.

1.2 Our contributions

We investigate the problem of designing commit-and-prove zero-knowledge systems for set membership and non-membership that can be used in a modular way and *efficiently* composed with other zero-knowledge proof systems for potentially arbitrary relations. Our main results are the following.

First, building upon the view of recent works on composable proofs [1, 18], we define a formal framework for commit-and-prove zkSNARKs (CP-SNARKs) for set (non-)membership. The main application of this framework is a compiler that, given a CP-SNARK $\mathsf{CP}_{\mathsf{mem}}$ for set membership and any other CP-SNARK CP_R for a relation R, yields a CP-SNARK CP for the composed relation "$u \in S \wedge \exists \omega : R(u, \omega)$". As a further technical contribution, our framework extends the one in [18] in order to work with commitments from multiple schemes (including set commitments, e.g., accumulators).

Second, we propose new efficient constructions of CP-SNARKs for set membership and non-membership, in which elements of the accumulated set can be committed with a Pedersen commitment in a prime order group $\mathbb{G}_q$—a setting that, as argued before, is of practical relevance due to the widespread use of these commitments and of proof systems that operate on them. In more detail, we propose: four schemes (two for set membership and two for non-membership) that enjoy constant-size public parameters and are based on RSA accumulators for committing to sets, and a scheme over pairings that has public parameters linear in the size of the set, but where the set can remain hidden.

Finally, we implement our solutions in a software library and experimentally evaluate their performance.

Like the recent works [1, 18], our work can be seen as showing yet another setting—set membership—where the efficiency of SNARKs can benefit from a modular design.

1.3 RSA-based constructions

Our first scheme, a CP-SNARK for set membership based on RSA accumulators, supports a large domain for the set of accumulated elements, represented by binary strings of a given length η. Our second scheme, also based on RSA accumulators, supports elements that are

[4] https://filecoin.io.

[5] More specifically: the elements of a set need to be prime numbers in a range (A, B) such that $q/2 > A^2 - 1 > B \cdot 2^{2\lambda_{st}+2}$. If aiming at 128 bits of security level one can meet this constraint by choosing for example $A = 2^{259}$, $B = 2^{260}$ and $q > 2^{519}$.

prime numbers of exactly μ bits (for a given μ). Neither scheme requires an a-priori bound on the cardinality of the set. Both schemes improve the proof-of-knowledge protocol by Camenisch and Lysyanskaya [16]: (i) we can work with a prime order group $\mathbb{G}_q$ of "standard" size, e.g., 256 bits, whereas [16] needs a much larger $\mathbb{G}_q$ (see above). We note that the size of $\mathbb{G}_q$ affects not only the efficiency of the set membership protocol but also the efficiency of any other protocol that needs to interact with commitments to alleged set members; (ii) we can support flexible choices for the size of set elements. For instance, in the second scheme, we could work with primes of about 50 or 80 bits,[6] which in practice captures virtually unbounded sets and can make the accumulator operations 4–5$\times$ faster compared to using $\approx$ 256-bits primes as in [16].

Our main technical contribution here involves a new way to link a proof of membership for RSA accumulators to a Pedersen commitment in a prime order group, together with a careful analysis showing this can be secure under parameters *not requiring a larger prime order group* (as in [16]). See Sect. 4 for further details.

1.4 Pairing-based construction

Our pairing-based scheme for set membership supports set elements in $\mathbb{Z}_q$, where q is the order of bilinear groups, while the sets are arbitrary subsets of $\mathbb{Z}_q$ of cardinality less than a fixed a-priori bound n. This scheme has the disadvantage of having public parameters linear in n, but has other advantages in comparison to previous schemes with a similar limitation (and also in comparison to the RSA-based schemes above). First, the commitment to the set can be hiding and untrusted for the verifier, i.e., the set can be kept hidden and it is not needed to check the opening of the commitment to the set; this makes it composable with proof systems that could for example prove global properties on the set, i.e., that $P(S)$ holds. Second, the scheme works entirely in bilinear groups, i.e., no need of operating over RSA groups. The main technical contribution here is a technique to turn the EDRAX vector commitment [23] into an accumulator admitting efficient zero-knowledge membership proofs.

1.5 Extensions to set non-membership

We propose extensions of both our CP-SNARK framework and RSA constructions to deal with proving *set non-membership*, namely proving in zero-knowledge that $u \notin S$ with respect to a commitment $c(u)$ and a committed set S. Our two RSA-based schemes for non-membership have the same features as the analogous membership schemes mentioned above: the first scheme supports sets whose elements are strings of length η, the second one supports elements that are prime numbers of μ bits, and both work with elements committed using Pedersen in a prime order group and sets committed with RSA accumulators. A byproduct of sharing the same parameters is that we can easily compose the set-membership and non-membership schemes, via our framework, in order to prove statements like $u \in S_1 \wedge u \notin S_2$. Our technical contribution in the design of these schemes is a zero-knowledge protocol for non-membership witnesses of RSA accumulators that is linked to Pedersen commitments in prime order groups.

[6] When prime representation is suitable for the application, distinct primes can be generated without a hash fuction (e.g. by using sequential primes).

1.6 Implementation and experiments

We have implemented our RSA-based[7] schemes for membership and non-membership as a Rust library which is publicly available [28]. Our library is implemented in a modular fashion to work with any elliptic curve from *libzexe* [67] and Ristretto from *curve25519-dalek* [54]. This choice enables everyone to easily and efficiently combine our CP-SNARKs in a modular way with other CP-SNARKs implemented over these elliptic curves, such as Bulletproofs [13] and LegoGroth16 [18].

We evaluated our RSA-based constructions and compared them against highly optimized solutions based on Merkle Trees.[8] Our schemes achieve significantly better performance in proving time while slightly compromising on proof size and verification time. Our implementation is fast, yet we have not heavily optimized it and thus expect the results can be further improved.

Our solutions supporting sets of arbitrary elements achieve a proving time that is up to[9] 3.7× faster for set membership (309 ms vs. 1.14 s) and up to 7× faster for set non-membership (325 ms vs. 2.28 s).[10]

Our solutions where elements of the set are large prime numbers (i.e., of 252-bit size) offer even better results: our proving time is 4.5×–23.5× faster for membership and 6.8×–36× faster for non-membership (depending on the depth of the Merkle tree used in the comparison). We also show an optimization that, at the price of achieving computational (instead of statistical) zero-knowledge, is twice faster (see Sect. 7.4). This scenario can for example capture the case of sets made of hiding commitments that are prime numbers. In Sect. 8 we discuss how this can be relevant for a slight variant of the Zerocash protocol where commitments can be made prime numbers.

More details on the implementation and the benchmarks are available in Sect. 7.

1.7 Transparent instantiations

We generalize our building blocks for RSA groups to any hidden-order group (Appendix 4). By instantiating the latter with class groups and by using a transparent CP-NIZK such as Bulletproofs, we obtain variants of our RSA-based schemes with *transparent setup*. Class groups are more expensive than traditional RSA groups; in this setting we still obtain performance (proving time 12s; $|\Pi| = 6.4$ kB) outperforming other transparent solution for large Merkle trees, roughly 2^{64} leaves (see [79, Fig. 5] which summarizes performances of transparent SNARKs used to prove Merkle tree computations using SHA256 as hash). These potential gains come at the price of a relatively longer verification (compared to other solutions): 6.4 s.

1.8 Other related work

Ozdemir et al. [58] recently proposed a solution to scale operations on RSA accumulators inside a SNARK. In particular, their approach scales when these operations are *batched* (i.e.,

[7] For the implementation we focused on schemes where the public parameters do not depend on the set size; hence, we did not implement the pairing-based solutions.

[8] For our experiments we consider Merkle Trees using Pedersen Hash over the JubJub curve [49].

[9] We stress the proving time for our construction does not vary when the set grows. On the other hand this time varies for solutions based on Merkle trees.

[10] These ratios refer to a comparison against Interval Merkle Trees which require opening two paths to prove non-membership. When compared against Sparse Merkle Trees, our solutions show similar improvement ratios.

when proving membership of many elements at the same time); for example, they surpass a 2^{20}-large Merkle tree when proving batches of at least 600 elements. This approach is attractive in settings where we can delegate a *large* quantity of these checks to an untrusted server as there is a high constant proving cost. In contrast, our approach can achieve faster proving time than Merkle trees already for a single membership check. It is an interesting open problem to adapt our techniques for modular set (non-)membership for the case of batched membership while keeping the tested elements hidden.

1.9 Organization

We give basic definitions in Sect. 2. In Sect. 3 we formalize commit-and-prove zkSNARKs for set (non-)membership. We describe our main constructions based on RSA accumulators for set membership and non-membership respectively in Sects. 4 and 5. We describe our construction for set membership based on bilinear pairings in Sect. 6. Finally, in Sects. 7 and 8 we discuss our implementation, experiments and applications.

1.10 Recent developments

Here we mention recent developments in the area of zero-knowledge proof for set (non-)membership, following the conference version of this paper published in 2021 [8].

A closely related work is that of Campanelli et al. [19] who present zero-knowledge protocols for RSA Accumulators with which one can prove membership for any number of Pedersen-committed elements (a so-called 'batch proof'). That is the proofs of [19] are independent both of the size of the set and the number of elements proving membership for.

In the bilinear groups setting, Srinivasan et al. [70], among other improvements on the functionalities and security properties of the actual pairing-based accumulator, provide zero-knowledge (batch) proofs for membership and non-membership over the Nguyen accumulator [57].

Another relevant, rapidly developing, line of work has to do with succinct zero-knowledge lookup arguments. That is, given a committed *vector* of n elements, one proves that a number m of committed elements are all values of the vector in some hidden position, while retaining the elements secret. The proofs are succinct in both n and m. This line of work was initiated by the seminal work of Zapico et al. [74] followed by a number of works improving the prover's complexity [35, 42, 62, 75]. All these constructions work over bilinear groups.

Finally, Lipmaa and Parisella [53] (building on [24, 26]) construct succinct set (non-)membership NIZKs from falsifaible assumptions. That is, the objective of their work is constructing efficient NIZKs for set (non-)membership that can be proven secure in the standard model and assuming only falsifiable assumptions.

1.11 Publication note

This article is the long version of the homonymous paper that appeared in the proceedings of Financial Cryptography and Data Security 2021 [8]. This version additionally contains:

- The Sect. 1.10 on recent developments (subsequent to [8] works) in the area.
- The full definitional framework of CP-SNARKs for set (non-)membership (Sect. 3).
- The pairing-based construction of Sect. 6.
- Full security proofs of the RSA-based constructions (Sects. 4, 5).

– An experimental evaluation of our RSA-based protocols (Sect. 7).
– A (slightly) different variant of our non-membership protocol (Appendix 2).
– A discussion on how to extend our RSA-based protocols to work with any Hidden Order Group (Appendix 4).

2 Preliminaries

2.1 Notation

We denote the security parameter with $\lambda \in \mathbb{N}$ and its unary representation with 1^λ. Throughout the paper we assume that all the algorithms of the cryptographic schemes take as input 1^λ, which is thus omitted from the list of inputs. If D is a distribution, we denote by $x \leftarrow D$ the process of sampling x according to D. An ensemble $\mathcal{X} = \{X_\lambda\}_{\lambda\in\mathbb{N}}$ is a family of probability distributions over a family of domains $\mathcal{D} = \{D_\lambda\}_{\lambda\in\mathbb{N}}$, and we say that two ensembles $\mathcal{D} = \{D_\lambda\}_{\lambda\in\mathbb{N}}$ and $\mathcal{D}' = \{D'_\lambda\}_{\lambda\in\mathbb{N}}$ are statistically indistinguishable (denoted by $\mathcal{D} \approx_s \mathcal{D}'$) if $\frac{1}{2}\sum_x |D_\lambda(x) - D'_\lambda(x)| < \mathsf{negl}(\lambda)$. If $\mathcal{A} = \{\mathcal{A}_\lambda\}$ is a (possibly non-uniform) family of circuits and $\mathcal{D} = \{D_\lambda\}_{\lambda\in\mathbb{N}}$ is an ensemble, then we denote by $\mathcal{A}(\mathcal{D})$ the ensemble of the outputs of $\mathcal{A}_\lambda(x)$ when $x \leftarrow D_\lambda$. We say two ensembles $\mathcal{D} = \{D_\lambda\}_{\lambda\in\mathbb{N}}$ and $\mathcal{D}' = \{D'_\lambda\}_{\lambda\in\mathbb{N}}$ are computationally indistinguishable (denoted by $\mathcal{D} \approx_c \mathcal{D}'$) if for every non-uniform polynomial time distinguisher $\mathcal{A}$ we have $\mathcal{A}(\mathcal{D}) \approx_s \mathcal{A}(\mathcal{D}')$.

We use $[n]$ to denote the set of integers $\{1, \ldots, n\}$, and $[0, n]$ for $\{0, 1, \ldots, n\}$. We denote by $(u_j)_{j\in[\ell]}$ the tuple of elements $(u_1, \ldots, u_\ell)$.

We denote $\mathsf{Primes} := \{e \in \mathbb{N} : e \text{ is prime}\}$ the set of all positive integers $e > 1$ such that they do not have non-trivial (i.e. different than e and 1) factors. More specifically, given two positive integers $A, B > 0$ such that $A < B$, we denote with $\mathsf{Primes}(A, B)$ the subset of Primes of numbers lying in the interval (A, B), i.e., $\mathsf{Primes}(A, B) := \{e \in \mathbb{Z} : e \text{ is prime } \wedge\ A < e < B\}$. According to the well known prime number theorem $|\mathsf{Primes}(1, B)| = O\left(\frac{B}{\log B}\right)$ which results to $|\mathsf{Primes}(A, B)| = O\left(\frac{B}{\log B}\right) - O\left(\frac{A}{\log A}\right)$.

2.2 RSA groups

We say that $N = pq$ is an RSA modulus for some primes p, q, such that $|p| = |q|$. We further say that N is a strong RSA modulus if there are primes p', q' such that $p = 2p'+1$, $q = 2q'+1$. We call $\mathbb{Z}_N^*$ for an RSA modulus an RSA group. With $\phi : \mathbb{N} \to \mathbb{N}$ we denote the Euler's totient function, $\phi(N) := |\mathbb{Z}_N^*|$. In particular for RSA modulus $\phi(N) = (p-1)(q-1)$. An RSA Group generator $N \leftarrow\!\!\$\ \mathsf{GenSRSAmod}(1^\lambda)$ is a probabilistic algorithm that outputs a strong RSA modulus N of bit-length $\ell(\lambda)$ for an appropriate polynomial $\ell(\cdot)$.

For any N we denote by $\mathsf{QR}_N := \{Y : \exists X \in \mathbb{Z}_N^* \text{ such that } Y = X^2 \pmod N\}$, the set of all the quadratic residues modulo N. QR_N is a subgroup (and thus closed under multiplication) of $\mathbb{Z}_N^*$ with order $|\mathsf{QR}_N| = |\mathbb{Z}_N^*|/2$. In particular for a strong RSA modulus $|\mathsf{QR}_N| = \frac{4p'q'}{2} = 2p'q'$.

2.2.1 Computational assumptions in RSA groups

The most fundamental assumption for RSA groups is the factoring assumption which states that given an RSA modulus $N \leftarrow \mathsf{GenSRSAmod}(1^\lambda)$ it is hard to compute its factors p and q. We further recall the Discrete Logarithm and strong RSA [2] assumptions:

Definition 2.1 *(DLOG assumption for RSA groups)* We say that the *Discrete Logarithm* (DLOG) assumption holds for GenSRSAmod if for any PPT adversary $\mathcal{A}$:

$$\Pr\left[\begin{array}{c} G^{x'} = Y \pmod N \\ : \\ \end{array}\begin{array}{l} N \leftarrow \mathsf{GenSRSAmod}(1^\lambda) \\ G \leftarrow_\$ \mathbb{Z}_N^*; x \leftarrow_\$ \mathbb{Z} \\ Y \leftarrow G^x \pmod N \\ x' \leftarrow \mathcal{A}(\mathbb{Z}_N^*, G, Y) \end{array}\right] = \mathsf{negl}(\lambda).$$

Definition 2.2 *(Strong-RSA assumption* [2]*)* We say that the *strong RSA assumption* holds for GenSRSAmod if for any PPT adversary $\mathcal{A}$:

$$\Pr\left[\begin{array}{c} U^e = G \\ \wedge e \neq 1, -1 \\ \end{array}\begin{array}{l} N \leftarrow \mathsf{GenSRSAmod}(1^\lambda) \\ G \leftarrow_\$ \mathbb{Z}_N^* \\ (U, e) \leftarrow \mathcal{A}(\mathbb{Z}_N^*, G) \end{array}\right] = \mathsf{negl}(\lambda).$$

2.3 Non-interactive zero-knowledge (NIZK)

We recall the definition of zero-knowledge non-interactive arguments of knowledge (NIZKs, for short).

Definition 2.3 *(NIZK)* A NIZK for $\{\mathcal{R}_\lambda\}_{\lambda\in\mathbb{N}}$ is a tuple of three algorithms $\Pi = (\mathsf{KeyGen}, \mathsf{Prove}, \mathsf{VerProof})$ that work as follows and satisfy the notions of *completeness*, *knowledge soundness* and (composable) *zero-knowledge* defined below.

- $\mathsf{KeyGen}(R) \rightarrow (\mathsf{ek}, \mathsf{vk})$ takes the security parameter λ and a relation $R \in \mathcal{R}_\lambda$, and outputs a common reference string consisting of an evaluation and a verification key.
- $\mathsf{Prove}(\mathsf{ek}, x, w) \rightarrow \pi$ takes an evaluation key for a relation R, a statement x, and a witness w such that $R(x, w)$ holds, and returns a proof π.
- $\mathsf{VerProof}(\mathsf{vk}, x, \pi) \rightarrow b$ takes a verification key, a statement x, and either accepts ($b = 1$) or rejects ($b = 0$) the proof π.

Completeness For any $\lambda \in \mathbb{N}$, $R \in \mathcal{R}_\lambda$ and (x, w) such that $R(x, w)$, it holds $\Pr[(\mathsf{ek}, \mathsf{vk}) \leftarrow \mathsf{KeyGen}(R), \pi \leftarrow \mathsf{Prove}(\mathsf{ek}, x, w) : \mathsf{VerProof}(\mathsf{vk}, x, \pi) = 1] = 1$.
Knowledge soundness Let $\mathcal{RG}$ be a relation generator such that $\mathcal{RG}_\lambda \subseteq \mathcal{R}_\lambda$. Π has computational knowledge soundness for $\mathcal{RG}$ and auxiliary input distribution $\mathcal{Z}$, denoted $\mathsf{KSND}(\mathcal{RG}, \mathcal{Z})$ for brevity, if for every (non-uniform) efficient adversary $\mathcal{A}$ there exists a (non-uniform) efficient extractor $\mathcal{E}$ such that $\Pr[\mathsf{Game}^{\mathsf{KSND}}_{\mathcal{RG},\mathcal{Z},\mathcal{A},\mathcal{E}} = 1] = \mathsf{negl}$. We say that Π is knowledge sound if there exists benign $\mathcal{RG}$ and $\mathcal{Z}$ such that Π is $\mathsf{KSND}(\mathcal{RG}, \mathcal{Z})$.

$\mathsf{Game}^{\mathsf{KSND}}_{\mathcal{RG},\mathcal{Z},\mathcal{A},\mathcal{E}} \rightarrow b$

$(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)$; $\mathsf{crs} := (\mathsf{ek}, \mathsf{vk}) \leftarrow \mathsf{KeyGen}(R)$
$\mathsf{aux}_Z \leftarrow \mathcal{Z}(R, \mathsf{aux}_R, \mathsf{crs})$; $(x, \pi) \leftarrow \mathcal{A}(R, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$
$w \leftarrow \mathcal{E}(R, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$; $b = \mathsf{VerProof}(\mathsf{vk}, x, \pi) \wedge \neg R(x, w)$

Composable zero-knowledge A scheme Π satisfies composable zero-knowledge for a relation generator $\mathcal{RG}$ if there exists a simulator $\mathcal{S} = (\mathcal{S}_{\mathsf{kg}}, \mathcal{S}_{\mathsf{prv}})$ such that both following conditions hold.

Keys indistinguishability For all adversaries $\mathcal{A}$

$$\Pr\left[\begin{array}{l}(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)\\ \mathsf{crs} \leftarrow \mathsf{KeyGen}(R)\\ \mathcal{A}(\mathsf{crs}, \mathsf{aux}_R) = 1\end{array}\right] \approx \Pr\left[\begin{array}{l}(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)\\ (\mathsf{crs}, \mathsf{td}_\mathsf{k}) \leftarrow \mathcal{S}_{\mathsf{kg}}(R)\\ \mathcal{A}(\mathsf{crs}, \mathsf{aux}_R) = 1\end{array}\right].$$

*P*roof indistinguishability For all adversaries $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$

$$\Pr\left[\begin{array}{l}(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)\\ (\mathsf{crs}, \mathsf{td}_\mathsf{k}) \leftarrow \mathcal{S}_{\mathsf{kg}}(R)\\ (x, w, \mathsf{st}) \leftarrow \mathcal{A}_1(\mathsf{crs}, \mathsf{aux}_R) \ : \ R(x, w)\\ \pi \leftarrow \mathsf{Prove}(\mathsf{ek}, x, w)\\ \mathcal{A}_2(\mathsf{st}, \pi) = 1\end{array}\right]$$

$$\approx \Pr\left[\begin{array}{l}(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)\\ (\mathsf{crs}, \mathsf{td}_\mathsf{k}) \leftarrow \mathcal{S}_{\mathsf{kg}}(R)\\ (x, w, \mathsf{st}) \leftarrow \mathcal{A}_1(\mathsf{crs}, \mathsf{aux}_R) \ : \ R(x, w)\\ \pi \leftarrow \mathcal{S}_{\mathsf{prv}}(\mathsf{crs}, \mathsf{td}_\mathsf{k}, x)\\ \mathcal{A}_2(\mathsf{st}, \pi) = 1\end{array}\right].$$

Definition 2.4 *(zkSNARKs)* A NIZK Π is called *zero-knowledge succinct non-interactive argument of knowledge* (zkSNARK) if Π is a NIZK as per Definition 2.3 enjoying an additional property, *succinctness*, i.e., if the running time of $\mathsf{VerProof}$ is $\mathsf{poly}(\lambda + |x| + \log|w|)$ and the proof size is $\mathsf{poly}(\lambda + \log|w|)$.

Remark 2.1 (On knowledge-soundness) In the NIZK definition above we use a non black-box notion of extractability. Although this is virtually necessary in the case of zkSNARKs [44], NIZKs can also satisfy stronger (black-box) notions of knowledge-soundness.

2.4 Type-based commitments

We recall the notion of Type-Based Commitment schemes introduced by Escala and Groth [36]. In brief, a Type-Based Commitment scheme is a normal commitment scheme with the difference that it allows one to commit to values from different domains. More specifically, the Commit algorithm (therefore the $\mathsf{VerCommit}$ algorithm also) depends on the domain of the input, while the commitment key remains the same. For example, as in the original motivation of [36], the committer can use the same scheme and key to commit to elements that may belong to two different groups $\mathbb{G}_1, \mathbb{G}_2$ or a field $\mathbb{Z}_p$. In our work we use type-based commitments. The main benefit of this formalization is that it can unify many commitment algorithms into one scheme. In our case this is useful to formalize the notion of commit-and-prove NIZKs that work with commitments from different groups and schemes.

More formally, a Type-Based Commitment is a tuple of algorithms $\mathsf{Com} = (\mathsf{Setup}, \mathsf{Commit}, \mathsf{VerCommit})$ that works as a Commitment scheme defined above with the difference that Commit and $\mathsf{VerCommit}$ algorithms take an extra input t that represent the type of u. All the possible types are included in the type space $\mathcal{T}$.[11]

Definition 2.5 A type-based commitment scheme for a set of types $\mathcal{T}$ is a tuple of algorithms $\mathsf{Com} = (\mathsf{Setup}, \mathsf{Commit}, \mathsf{VerCommit})$ that work as follows:

[11] Normally $\mathcal{T}$ is finite and includes a small number of type, e.g. $\mathcal{T} = \{\mathbb{G}_1, \mathbb{G}_2, \mathbb{Z}_p\}$.

- $\mathsf{Setup}(1^\lambda) \to \mathsf{ck}$ takes the security parameter and outputs a commitment key ck. This key includes $\forall \mathsf{t} \in \mathcal{T}$ descriptions of the input space $\mathcal{D}_\mathsf{t}$, commitment space $\mathcal{C}_\mathsf{t}$ and opening space $\mathcal{O}_\mathsf{t}$.
- $\mathsf{Commit}(\mathsf{ck}, \mathsf{t}, u) \to (c, o)$ takes the commitment key ck, the type t of the input and a value $u \in \mathcal{D}_\mathsf{t}$, and outputs a commitment c and an opening o.
- $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}, c, u, o) \to b$ takes as a type t, a commitment c, a value u and an opening o, and accepts ($b = 1$) or rejects ($b = 0$).

Furthermore, the security properties depend on the type, in the sense that binding and hiding should hold with respect to a certain type.

Definition 2.6 Let $\mathcal{T}$ be a set of types, and Com be a type-based commitment scheme for $\mathcal{T}$. Correctness, t-Type Binding and t-Type Hiding are defined as follows:
Correctness For all $\lambda \in \mathbb{N}$ and any input $(\mathsf{t}, u) \in (\mathcal{T}, \mathcal{D}_\mathsf{t})$ we have:

$$\Pr[\mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda), (c, o) \leftarrow \mathsf{Commit}(\mathsf{ck}, \mathsf{t}, u) : \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}, c, u, o) = 1] = 1.$$

t-*Type binding* Given $\mathsf{t} \in \mathcal{T}$, for every polynomial-time adversary $\mathcal{A}$:

$$\Pr\left[\begin{array}{l} \mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda) \\ (c, u, o, u', o') \leftarrow \mathcal{A}(\mathsf{ck}, \mathsf{t}) \end{array} : \begin{array}{r} u \neq u' \wedge \ \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}, c, u, o) = 1 \\ \wedge \ \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}, c, u', o') = 1 \end{array}\right] = \mathsf{negl}.$$

In case Com is t-type bidning for all $\mathsf{t} \in \mathcal{T}$ we will say that it is Binding.
t -*Type hiding* Given a $\mathsf{t} \in \mathcal{T}$, for $\mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda)$ and every pair of values $u, u' \in \mathcal{D}_\mathsf{t}$, the following two distributions are statistically close: $\mathsf{Commit}(\mathsf{ck}, \mathsf{t}, u) \approx \mathsf{Commit}(\mathsf{ck}, \mathsf{t}, u')$.
In case Com is t-Type Hiding for all $\mathsf{t} \in \mathcal{T}$ we say it is Hiding.

Composing type-based commitments For simplicity we now define an operator that allows to compose type-based commitment schemes in a natural way.

Definition 2.7 Let C and C' be two commitment schemes respectively for (disjoint) sets of types $\mathcal{T}$ and $\mathcal{T}'$. Then we denote by $\mathsf{C} \bullet \mathsf{C}'$ the commitment scheme $\bar{\mathsf{C}}$ for $\mathcal{T} \cup \mathcal{T}'$ such as:

- $\bar{\mathsf{C}}.\mathsf{Setup}(\mathsf{secpar}, \mathsf{secpar}') \to \overline{\mathsf{ck}}$: compute $\mathsf{ck} \leftarrow \mathsf{C}.\mathsf{Setup}(\mathsf{secpar})$ and $\mathsf{ck}' \leftarrow \mathsf{C}'.\mathsf{Setup}(\mathsf{secpar}')$; $\overline{\mathsf{ck}} := (\mathsf{ck}, \mathsf{ck}')$.
- $\bar{\mathsf{C}}.\mathsf{Commit}(\overline{\mathsf{ck}} := (\mathsf{ck}, \mathsf{ck}'), \mathsf{t}, u)$: If $t \in \mathcal{T}$ then output $\mathsf{C}.\mathsf{Commit}(\mathsf{ck}, \mathsf{t}, u)$; otherwise return $\mathsf{C}'.\mathsf{Commit}(\mathsf{ck}', \mathsf{t}, u)$.
- $\bar{\mathsf{C}}.\mathsf{VerCommit}(\overline{\mathsf{ck}} := (\mathsf{ck}, \mathsf{ck}'), \mathsf{t}, c, u, o)$: If $t \in \mathcal{T}$ then return $\mathsf{C}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}, c, u, o)$; otherwise return $\mathsf{C}'.\mathsf{VerCommit}(\mathsf{ck}', \mathsf{t}, c, u, o)$.

The following property of $\bullet$ follows immediately from its definition.

Lemma 2.1 *Let* C *and* C' *be two commitment schemes with disjoint sets of types. For all types* t *if* C *or* C' *is* t*-hiding (resp.* t*-binding) then* $\mathsf{C} \bullet \mathsf{C}'$ *is* t*-hiding (resp.* t*-binding).*

Remark 2.2 We observe that a standard non type-based commitment scheme with input space $\mathcal{D}$ induces directly a type-based commitment scheme with the same input space and a type we denote by $\mathbb{T}[\mathcal{D}]$.

2.5 Commit-and-prove NIZKs

We give the definition of *commit-and-prove NIZKs* (CP-NIZKs). We start from the definition given in [7, 18] and we extend it to type-based commitments. The main benefit of such

extension is that we can formalize CP-NIZKs working with commitments over different domains. In a nutshell, a CP-NIZK is a NIZK that can prove knowledge of (x, w) such that $R(x, w)$ holds with respect to a witness $w = (u, \omega)$ such that u opens a commitment c_u. As done in [18], we explicitly considers the input domain $\mathcal{D}_u$ at a more fine grained-level splitting it over ℓ subdomains. We call them *commitment slots* as each of the $\mathcal{D}_i$-s intuitively corresponds to a committed element.[12] The description of the splitting is assumed part of R's description.

In the remainder of this work we use the following shortcut definition. If C is a type-based commitment scheme over set of types $\mathcal{T}$, we say that a relation R over $(\mathcal{D}_1 \times \cdots \times \mathcal{D}_\ell)$ is $\mathcal{T}$-compatible if for all $j \in [\ell]$ it holds that $\mathbb{T}[\mathcal{D}_j] \in \mathcal{T}$. We say a relation family $\mathcal{R}$ is $\mathcal{T}$-compatible if every R in $\mathcal{R}$ is $\mathcal{T}$-compatible; a relation generator $\mathcal{RG}$ is $\mathcal{T}$-compatible if $\mathsf{Range}(\mathcal{RG})$ is $\mathcal{T}$-compatible.

Definition 2.8 *(CP-NIZKs [18])* Let $\{\mathcal{R}_\lambda\}_{\lambda\in\mathbb{N}}$ be a family of relations R over $\mathcal{D}_x \times \mathcal{D}_u \times \mathcal{D}_\omega$ such that $\mathcal{D}_u$ splits over ℓ arbitrary domains $(\mathcal{D}_1 \times \cdots \times \mathcal{D}_\ell)$ for some arity parameter $\ell \geq 1$. Let $\mathsf{C} = (\mathsf{Setup}, \mathsf{Commit}, \mathsf{VerCommit})$ be a commitment scheme (as per Definition 2.5) over set of types $\mathcal{T}$ such that $\{\mathcal{R}_\lambda\}_{\lambda\in\mathbb{N}}$ is $\mathcal{T}$-compatible.

A commit and prove NIZK for C and $\{\mathcal{R}_\lambda\}_{\lambda\in\mathbb{N}}$ is a NIZK for a family of relations $\{\mathcal{R}^{\mathsf{C}}_\lambda\}_{\lambda\in\mathbb{N}}$ such that:

- every $\mathbf{R} \in \mathcal{R}^{\mathsf{C}}$ is represented by a pair (ck, R) where $\mathsf{ck} \in \mathsf{C.Setup}(1^\lambda)$ and $R \in \mathcal{R}_\lambda$;
- $\mathbf{R}$ is over pairs $(\mathbf{x}, \mathbf{w})$ where the statement is $\mathbf{x} := (x, (c_j)_{j\in[\ell]}) \in \mathcal{D}_x \times \mathcal{C}^\ell$, the witness is $\mathbf{w} := ((u_j)_{j\in[\ell]}, (o_j)_{j\in[\ell]}, \omega) \in \mathcal{D}_1 \times \cdots \times \mathcal{D}_\ell \times \mathcal{O}^\ell \times \mathcal{D}_\omega$, and the relation $\mathbf{R}$ holds iff

$$\bigwedge\nolimits_{j\in[\ell]} \mathsf{VerCommit}(\mathsf{ck}, \mathbb{T}[\mathcal{D}_j], c_j, u_j, o_j) = 1 \wedge R(x, (u_j)_{j\in[\ell]}, \omega) = 1.$$

We denote knowledge soundness of a CP-NIZK for commitment scheme C and relation and auxiliary input generators $\mathcal{RG}$ and $\mathcal{Z}$ as $\mathsf{CP\text{-}KSND}(\mathsf{C}, \mathcal{RG}, \mathcal{Z})$.

We denote a CP-NIZK as a tuple of algorithms $\mathsf{CP} = (\mathsf{KeyGen}, \mathsf{Prove}, \mathsf{VerProof})$. For ease of exposition, in our constructions we adopt the following explicit syntax for CP's algorithms.

- $\mathsf{KeyGen}(\mathsf{ck}, R) \to \mathsf{crs} := (\mathsf{ek}, \mathsf{vk})$
- $\mathsf{Prove}(\mathsf{ek}, x, (c_j)_{j\in[\ell]}, (u_j)_{j\in[\ell]}, (o_j)_{j\in[\ell]}, \omega) \to \pi$
- $\mathsf{VerProof}(\mathsf{vk}, x, (c_j)_{j\in[\ell]}, \pi) \to b \in \{0, 1\}$

2.6 Commit-and-prove NIZKs with partial opening

We now define a variant of commit-and-prove NIZKs with a weaker notion of knowledge-soundness. In particular we consider the case where part of the committed input is not assumed to be extractable (or hidden),[13] i.e., such input is assumed to be opened by the adversary. This models scenarios where we do not require this element to be input of the verification algorithm (the verifier can directly use a digest to it).

The motivation to define and use this notion is twofold. First, in some constructions commitments on sets are compressing but not knowledge-extractable. Second, in many applications this definition is sufficient since the set is public (e.g., the set contain the valid coins).

[12] Each of the "open" elements in the $\mathcal{D}_i$-s (together with any auxiliary opening information) should also be thought of as the witness to the relation as we require them to be extractable. On the other hand, the commitments themselves are part of the public input.

[13] This is reminiscent of the soundness notions considered in [39].

The definition below is limited to a setting where the adversary opens only one input in this fashion.[14] We will assume, as a convention, that in a scheme with partial opening this special input is always the first committed input of the relation, i.e. the one denoted by u_1 and corresponding to $\mathcal{D}_1$. We note that the commitment to u_1 does not require hiding for zero-knowledge to hold.

Definition 2.9 *(CP-NIZK with partial opening)* A commit and prove NIZK with partial opening for C and $\{\mathcal{R}_\lambda\}_{\lambda\in\mathbb{N}}$ is a NIZK for a family of relations $\{\mathcal{R}_\lambda^{\mathsf{C}}\}_{\lambda\in\mathbb{N}}$ (defined as in Definition 2.8) such that the property of knowledge soundness is replaced by *knowledge soundness with partial opening* below.

Knowledge soundness with partial opening Let $\mathcal{RG}$ be a relation generator such that $\mathcal{RG}_\lambda \subseteq \mathcal{R}_\lambda$. Π has knowledge soundness with partial opening for C, $\mathcal{RG}$ and auxiliary input distribution $\mathcal{Z}$, denoted $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \mathcal{RG}, \mathcal{Z})$ for brevity, if for every (non-uniform) efficient adversary $\mathcal{A}$ there exists a (non-uniform) efficient extractor $\mathcal{E}$ such that $\Pr[\mathsf{Game}^{\mathsf{CP\text{-}poKSND}}_{\mathsf{C},\mathcal{RG},\mathcal{Z},\mathcal{A},\mathcal{E}} = 1] = \mathsf{negl}$. We say that Π is knowledge sound for C if there exists benign $\mathcal{RG}$ and $\mathcal{Z}$ such that Π is $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \mathcal{RG}, \mathcal{Z})$.[15]

$\mathsf{Game}^{\mathsf{CP\text{-}poKSND}}_{\mathsf{C},\mathcal{RG},\mathcal{Z},\mathcal{A},\mathcal{E}} \to b$

$\mathsf{ck} \leftarrow \mathsf{C.Setup}(1^\lambda);\ (R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda);\ \mathbf{R} := (\mathsf{ck}, R)$

$\mathsf{crs} := (\mathsf{ek}, \mathsf{vk}) \leftarrow \mathsf{KeyGen}(\mathbf{R})$

$\mathsf{aux}_Z \leftarrow \mathcal{Z}(\mathbf{R}, \mathsf{aux}_R, \mathsf{crs})$

$\left(x, (c_j)_{j\in[\ell]}, u_1, o_1, \pi\right) \leftarrow \mathcal{A}(\mathbf{R}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$

$\left((u_j)_{j\in[\ell]}, (o_j)_{j\in[\ell]}, \omega\right) \leftarrow \mathcal{E}(\mathbf{R}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$

$b = \mathsf{VerProof}(\mathsf{vk}, x, (c_j)_{j\in[\ell]}, \pi) \wedge$

$\neg\left(\bigwedge_{j\in[\ell]} \mathsf{C.VerCommit}(\mathsf{ck}, \mathbb{T}[\mathcal{D}_j], c_j, u_j, o_j) = 1 \wedge R(x, (u_j)_{j\in[\ell]}, \omega) = 1\right)$

Remark 2.3 (On weaker ZK in the context of partial opening) The notion of zero-knowledge for CP-NIZKs with partial opening that is implied by our definition above implies that the simulator does not have access to the opening of the first input (as it is the case in zero-knowledge for CP-NIZKs in general). Since this first commitment is opened, in principle one could also consider and define a weaker notion of zero-knowledge where the simulator has access to the first opened input. We leave it as an open problem to investigate if it can be of any interest.

Remark 2.4 (Full extractability) If a CP-NIZK has an empty input u_1 opened by the adversary in the game above, then we say that it is *fully extractable*. This roughly corresponds to the notion of knowledge soundness in Definition 2.3.

[14] We can easily generalize the notion for an adversary opening an arbitrary subset of the committed inputs.

[15] We point out that, although in the game below we make explicit the commitment opening in the relation, this is essentially the same notion of knowledge soundness as in CP-NIZKs (i.e. Definition 2.3) where the only tweak is that the adversary gives explicitly the first input in the commitment slot. We make commitments explicit hoping for the definition to be clearer. This is, however, in contrast to the definition of CP-NIZKs where the commitment opening is completely abstracted away inside the relation.

$\mathsf{Aux}^{\mathcal{RG}}(1^\lambda)$:

$(R_1, \mathsf{aux}_R^{(1)}) \leftarrow \mathcal{RG}_1(1^\lambda)$

$(R_2, \mathsf{aux}_R^{(2)}) \leftarrow \mathcal{RG}_2(1^\lambda)$

return $(R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}}$

$\mathcal{RG}^*(1^\lambda)$:

$(R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}} \leftarrow \mathsf{Aux}^{\mathcal{RG}}(1^\lambda)$

return $(R^\wedge_{R_1,R_2}, (\mathsf{aux}_R^{(b)})_{b\in\{1,2\}})$

$\overline{\mathcal{RG}}_b(1^\lambda)$:

$(R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}}$

$\leftarrow \mathsf{Aux}^{\mathcal{RG}}(1^\lambda)$

return $(R_b, \overline{\mathsf{aux}}_R^{(b)}$

$:= (R_{3-b}, (\mathsf{aux}_R^{(b)})_{b\in\{1,2\}}))$

$\mathsf{Aux}^{\mathcal{Z}}(\mathsf{ck}, (\mathsf{crs}_b, R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}})$:

$\mathsf{aux}_Z^{(1)} \leftarrow \mathcal{Z}_1(\mathsf{ck}, R_1, \mathsf{crs}_1, \mathsf{aux}_R^{(1)})$

$\mathsf{aux}_Z^{(2)} \leftarrow \mathcal{Z}_2(\mathsf{ck}, R_2, \mathsf{crs}_2, \mathsf{aux}_R^{(2)})$

return $(\mathsf{aux}_Z^{(b)})_{b\in\{1,2\}}$

$\mathcal{Z}^*((\mathsf{ck}, R^\wedge_{R_1,R_2}), (\mathsf{ek}^*, \mathsf{vk}^*), (\mathsf{aux}_R, \mathsf{aux}'_R))$:

$(\mathsf{aux}_Z^{(b)})_{b\in\{1,2\}}$

$\leftarrow \mathsf{Aux}^{\mathcal{Z}}(\mathsf{ck}, (\mathsf{crs}_b, R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}})$

return $(\mathsf{aux}_Z^{(b)})_{b\in\{1,2\}}$

$\overline{\mathcal{Z}}_b(\mathsf{ck}, R_b, \mathsf{crs}_b, \overline{\mathsf{aux}}_R^{(b)})$:

Parse $\overline{\mathsf{aux}}_R$ as $(R_{3-b}, (\mathsf{aux}_R^{(b)})_{b\in\{1,2\}})$

$\mathsf{crs}_{3-b} \leftarrow \mathsf{CP}_{3-b}.\mathsf{KeyGen}(\mathsf{ck}, R_{3-b})$

$(\mathsf{aux}_Z^{(b)})_{b\in\{1,2\}} \leftarrow \mathsf{Aux}^{\mathcal{Z}}(\mathsf{ck}, \ldots$

$\ldots, (\mathsf{crs}_b, R_b, \mathsf{aux}_R^{(b)})_{b\in\{1,2\}})$

$\overline{\mathsf{aux}}_Z^{(b)} := (\mathsf{crs}_{3-b}, (\mathsf{aux}_Z^{(b)})_{b\in\{1,2\}})$

return $\overline{\mathsf{aux}}_Z^{(b)}$

Fig. 1 Relation and auxiliary input generators for AND composition construction

2.6.1 Composition properties of commit-and-prove schemes

In [18], Campanelli et al. show a compiler for composing commit-and-prove schemes that work for the same commitment scheme in order to obtain CP systems for conjunction of relations. In this section we generalize their results to the case of typed relations and type-based commitments. This generalization in particular can model the composition of CP-NIZKs that work with different commitments, as is the case in our constructions for set membership in which one has a commitment to a set and a commitment to an element.

We begin by introducing the following compact notation for an augmented relation generator.

Definition 2.10 *(Augmented relation generator)* Let $\mathcal{RG}$ be a relation generator and $\mathcal{F}(1^\lambda)$ an algorithm taking as input a security parameter. Then we denote by $\mathcal{RG}[\mathcal{F}]$ the relation generator returning $(R, (\mathsf{aux}_R, \mathsf{out}_\mathcal{F}))$ where $\mathsf{out}_\mathcal{F} \leftarrow \mathcal{F}(1^\lambda)$ and $(R, \mathsf{aux}_R) \leftarrow \mathcal{RG}(1^\lambda)$.

The next lemma states that we can (with certain restrictions) trivially extend a CP-NIZK for commitment scheme C to an extended commitment scheme $\mathsf{C} \bullet \mathsf{C}'$.

Lemma 2.2 (Extending to commitment composition) *Let* C, C' *be commitment schemes defined over disjoint type sets* $\mathcal{T}$ *and* $\mathcal{T}'$*. If CP is* $\mathsf{CP}\text{-}\mathsf{poKSND}(\mathsf{C}, \mathcal{RG}[\mathsf{C.Setup}], \mathcal{Z})$ *for some relation and auxiliary input generators* $\mathcal{RG}, \mathcal{Z}$*. Then CP is* $\mathsf{CP}\text{-}\mathsf{poKSND}(\mathsf{C} \bullet \mathsf{C}', \mathcal{RG}[\mathsf{C.Setup}], \mathcal{Z})$ *if* $\mathcal{RG}$ *is* $\mathcal{T}$*-compatible.*

We now define relation generators and auxiliary input generators for our composition constructions.

The following lemma shows how we can compose CP-NIZKs even when one of them is fully extractable but the other is not. We are interested in the conjunction $R^\wedge_{asym}$ of relations

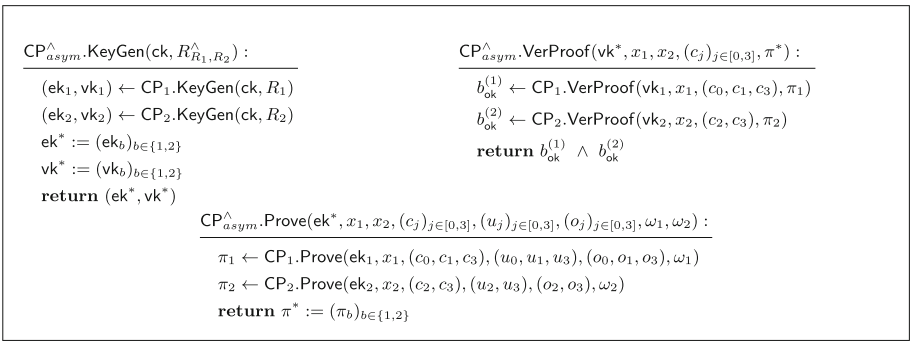


Fig. 2 CP-NIZK construction for AND composition (asymmetric case)

of type $R_1(x_1, (u_0, u_1, u_3), \omega_1)$ and $R_2(x_2, (u_2, u_3), \omega_2)$ where

$$R^{\wedge}_{asym}(x_1, x_2, (u_0, u_1, u_2, u_3), \omega_1, \omega_2) := R_1(x_1, (u_0, u_1, u_3), \omega_1) \wedge R_2(x_2, (u_2, u_3), \omega_2).$$

Lemma 2.3 (Composing conjunctions (with asymmetric extractability)) *Let* C *be a computationally binding commitment scheme. If* CP_1 *is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \overline{\mathcal{RG}}_1, \overline{\mathcal{Z}}_1)$ *and* CP_2 *is* $\mathsf{KSND}(\mathsf{C}, \overline{\mathcal{RG}}_2, \overline{\mathcal{Z}}_2)$ *(where* $\overline{\mathcal{RG}}_b, \overline{\mathcal{Z}}_b$ *are defined in terms of* $\mathcal{RG}_b, \mathcal{Z}_b$ *in Fig. 1 for* $b \in \{1, 2\}$*), then the scheme* $\mathsf{CP}^{\wedge}_{asym}$ *in Fig. 2 is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \mathcal{RG}^*, \mathcal{Z}^*)$ *where* $\mathcal{RG}^*, \mathcal{Z}^*$ *are as defined in Fig. 1.*

The following lemma is a symmetric variant of Lemma 2.3, i.e. the CP-NIZKs we are composing are both secure over the same commitment scheme and support partial opening, that is they both handle relations with and adversarially open input u_0. This time we are interested in the conjunction $R^{\wedge}_{sym}$ of relations of type $R_1(x_1, (u_0, u_1, u_3), \omega_1)$ and $R_2(x_2, (u_0, u_2, u_3), \omega_2)$ where

$$\begin{aligned} R^{\wedge}_{sym}(x_1, x_2, (u_0, u_1, u_2, u_3), \omega_1, \omega_2) := {} & R_1(x_1, (u_0, u_1, u_3), \omega_1) \\ & \wedge R_2(x_2, (u_0, u_2, u_3), \omega_2). \end{aligned}$$

Lemma 2.4 (Composing conjunctions (symmetric case)) *Let* C *be a (type-based) computationally binding commitment scheme. If* CP_b *is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \overline{\mathcal{RG}}_b, \overline{\mathcal{Z}}_b)$ *(where* $\overline{\mathcal{RG}}_b, \overline{\mathcal{Z}}_b$ *are defined in terms of* $\mathcal{RG}_b, \mathcal{Z}_b$ *in Fig. 1) for* $b \in \{1, 2\}$*, then the scheme* $\mathsf{CP}^{\wedge}_{sym}$ *in Fig. 3 is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}, \mathcal{RG}^*, \mathcal{Z}^*)$ *where* $\mathcal{RG}^*, \mathcal{Z}^*$ *are as defined in Fig. 1.*

3 CP-SNARKs for set membership (and non-membership)

In this section we discuss a specialization of CP-SNARKs for the specific NP relation that models membership (resp. non-membership) of an element in a set, formally defined below.

3.1 Set membership relations

Let $\mathcal{D}_{\mathsf{elm}}$ be some domain for set elements, and let $\mathcal{D}_{\mathsf{set}} \subseteq 2^{\mathcal{D}_{\mathsf{elm}}}$ be a set of possible sets over $\mathcal{D}_u$. We define the set membership relation $R_{\mathsf{mem}} : \mathcal{D}_{\mathsf{elm}} \times \mathcal{D}_{\mathsf{set}}$ as

$$R_{\mathsf{mem}}(U, u) = 1 \iff u \in U.$$

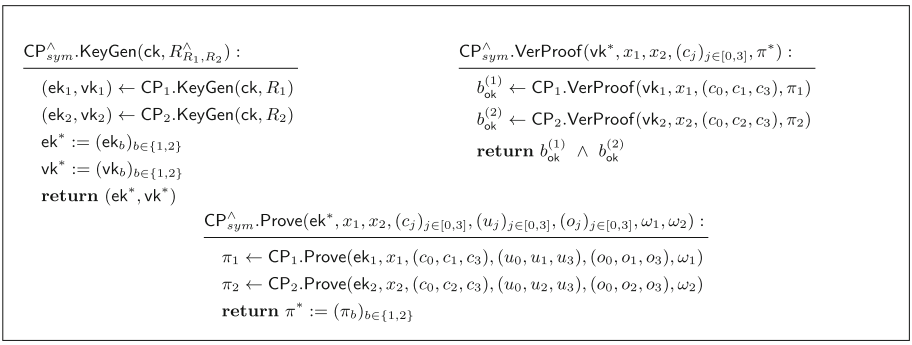


Fig. 3 CP-NIZK construction for AND composition (symmetric case)

This is the fundamental relation that we deal with in the rest of this work.

The non-membership relation $R_{\mathsf{nmem}} : \mathcal{D}_{\mathsf{elm}} \times \mathcal{D}_{\mathsf{set}}$ can be defined analogously as

$$R_{\mathsf{nmem}}(U, u) = 1 \iff u \notin U.$$

3.2 CP-SNARKs for set membership

Intuitively, a commit-and-prove SNARK for set membership allows one to commit to a set U and to an element u, and then to prove in zero-knowledge that $R_{\mathsf{mem}}(U, u) = 1$. More formally, let $R_{\mathsf{mem}} : \mathcal{D}_{\mathsf{elm}} \times \mathcal{D}_{\mathsf{set}}$ be a set membership relation as defined above where $\mathbb{T}[\mathcal{D}_{\mathsf{elm}}] = \mathsf{t}_{\mathsf{elm}}$ and $\mathbb{T}[\mathcal{D}_{\mathsf{set}}] = \mathsf{t}_{\mathsf{set}}$, and let $\mathsf{Com}_{\mathsf{S}\cup\mathsf{elm}}$ be a type-based commitment scheme for $\mathcal{T}$ such that $\mathsf{t}_{\mathsf{set}}, \mathsf{t}_{\mathsf{elm}} \in \mathcal{T}$. Basically, $\mathsf{Com}_{\mathsf{S}\cup\mathsf{elm}}$ allows one to either commit an element of $\mathcal{D}_{\mathsf{elm}}$ or to a set of values of $\mathcal{D}_{\mathsf{elm}}$. Then a CP-SNARK for set membership is a CP-SNARK for the family of relations $\{\mathcal{R}^{\mathsf{mem}}_{\lambda}\}$ and a type-based commitment scheme $\mathsf{Com}_{\mathsf{S}\cup\mathsf{elm}}$. It is deduced from definition 2.8 that this is a zkSNARK for the relation:

$$\mathbf{R} = (ck, R_{\mathsf{mem}}) \text{ over}$$
$$(\boldsymbol{x}, \boldsymbol{w}) = ((x, c), (u, o, \omega)) := ((\varnothing, (c_U, c_u)), ((U, u), (o_U, o_u), \varnothing)),$$

such that $\mathbf{R}$ holds iff:

$$R_{\mathsf{mem}}(U, u) = 1 \wedge \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathsf{set}}, c_U, U, o_U) = 1$$
$$\wedge \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathsf{elm}}, c_u, u, o_u) = 1.$$

A commit-and-prove version of R_{nmem} can be defined as a natural variant of the relation above.

Notice that for the relation R_{mem} it is relevant for the proof system to be succinct so that proofs can be at most polylogarithmic (or constant) in the the size of the set (that is part of the witness). This is why for set membership we are mostly interested in designing CP-SNARKs.

Proving arbitrary relations involving set (non-)membership

As discussed in the introduction, a primary motivation of proving set membership in zero-knowledge is to prove additional properties about an alleged set member. In order to make our CP-SNARK for set membership a reusable gadget, we discuss a generic and simple method for composing CP-SNARKs for set membership (with partial opening) with other

CP-SNARKs (with full extractability) for arbitrary relations. More formally, let R_{mem} be the set membership relation over pairs $(U, u) \in \mathbb{X} \times \mathcal{D}_u$ as R be an arbitrary relation over pairs (u, ω), then we define as R^* the relation:

$$R^*(U, u, \omega) := R_{\mathsf{mem}}(U, u) \wedge R(u, \omega).$$

The next corollary (direct consequence of Lemmas 2.2, 2.3) states we can straightforwardly compose a CP-SNARK for set membership with a CP-SNARK for an arbitrary relation on elements of the set.

Corollary 3.1 (Extending relations with set membership) *Let* $\mathsf{C}_{\mathsf{S}}, \mathsf{C}_u$ *be two computationally binding commitment schemes defined over disjoint type sets* $\mathcal{T}_{\mathsf{S}}$ *and* $\mathcal{T}_u$. *Let* $\mathsf{CP}_{\mathsf{mem}}, \mathsf{CP}_u$ *be two CP-SNARKs and* $R_{\mathsf{mem}}, \mathcal{RG}_u$ *(resp.* $\mathcal{Z}_{\mathsf{mem}}, \mathcal{Z}_u$*) be two relation (resp. auxiliary input) generators. If* $\mathsf{CP}_{\mathsf{mem}}$ *is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}_{\mathsf{S}} \bullet \mathsf{C}_u, R_{\mathsf{mem}}, \mathcal{Z}_{\mathsf{mem}})$ *and* CP_u *is* $\mathsf{KSND}(\mathsf{C}_u, \mathcal{RG}_u, \mathcal{Z}_u)$ *then there exists a* CP^* *that is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}_{\mathsf{S}} \bullet \mathsf{C}_u, \mathcal{RG}^*, \mathcal{Z}^*)$ *where* $\mathcal{RG}^*, \mathcal{Z}^*$ *are as defined in Fig. 1.*

In a similar fashion, we can combine an arbitrary relation R with the relation for non-membership obtaining relation $\bar{R}^*$ defined as:

$$\bar{R}^*(U, u, \omega) := R_{\mathsf{nmem}}(U, u) \wedge R(u, \omega).$$

The next corollary states we can straightforwardly compose a CP-SNARK for set non-membership with a CP-SNARK for an arbitrary relation on elements in the universe of the set.

Corollary 3.2 (Extending relations with set non-membership) *Let* $\mathsf{C}_{\mathsf{S}}, \mathsf{C}_u$ *be two computationally binding commitment schemes defined over disjoint type sets* $\mathcal{T}_{\mathsf{S}}$ *and* $\mathcal{T}_u$. *Let* $\mathsf{CP}_{\mathsf{nmem}}, \mathsf{CP}_u$ *be two CP-SNARKs and* $R_{\mathsf{nmem}}, \mathcal{RG}_u$ *(resp.* $\mathcal{Z}_{\mathsf{nmem}}, \mathcal{Z}_u$*) be two relation (resp. auxiliary input) generators. If* $\mathsf{CP}_{\mathsf{nmem}}$ *is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}_{\mathsf{S}} \bullet \mathsf{C}_u, R_{\mathsf{nmem}}, \mathcal{Z}_{\mathsf{nmem}})$ *and* CP_u *is* $\mathsf{KSND}(\mathsf{C}_u, \mathcal{RG}_u, \mathcal{Z}_u)$ *then there exists a* CP^* *that is* $\mathsf{CP\text{-}poKSND}(\mathsf{C}_{\mathsf{S}} \bullet \mathsf{C}_u, \mathcal{RG}^*, \mathcal{Z}^*)$ *where* $\mathcal{RG}^*, \mathcal{Z}^*$ *are as defined in Fig. 1.*

3.2.1 CP-SNARKs for set membership from accumulators with proofs of knowledge

As discussed in the introduction, CP-SNARKs for set membership are simply a different lens through which we can approach accumulators that have a protocol for proving in zero-knowledge that a committed value is in the accumulator (i.e., it is in the set succinctly represented by the accumulator). To strengthen this intuition in Appendix 2 we formally show that a CP-SNARK for set membership can be constructed from an accumulator scheme that has a zero-knowledge proof for committed values. This allows us to capture existing schemes such as [16, 57].

4 A CP-SNARK for set membership with short parameters

In this section we describe CP-SNARKs for set membership in which the elements of the sets can be committed using a Pedersen commitment scheme defined in a prime order group, and the sets are committed using an RSA accumulator. The advantage of having elements committed with Pedersen in a prime order group is that our CP-SNARKs can be composed with any other CP-SNARK for Pedersen commitments and relations R that take set elements

as inputs. The advantage of committing to sets using RSA accumulators is instead that the public parameters (i.e., the CRS) of the CP-SNARKs presented in this section are *short*, virtually independent of the size of the sets. Since RSA accumulators are not extractable commitments, the CP-SNARKs presented here are secure in a model where the commitment to the set is assumed to be checked at least once, namely they are knowledge-sound with partial opening of the set commitment.

A bit more in detail, we propose two CP-SNARKs. Our first scheme, called $\mathsf{MemCP}_{\mathsf{RSA}}$, works for set elements that are arbitrary strings of length η, i.e., $\mathcal{D}_{\mathsf{elm}} = \{0,1\}^{\eta}$, and for sets that are any subset of $\mathcal{D}_{\mathsf{elm}}$, i.e., $\mathcal{D}_{\mathsf{set}} = 2^{\mathcal{D}_{\mathsf{elm}}}$. Our second scheme, $\mathsf{MemCP}_{\mathsf{RSAPrm}}$, instead works for set elements that are prime numbers of exactly μ bits, and for sets that are any subset of such prime numbers. This second scheme is a simplified variant of the first one that requires more structure on the set elements (they must be prime numbers) but in exchange of that offers better efficiency. So it is preferable in those applications that can work with prime representatives.

4.1 An high-level overview of our constructions

We provide the main idea behind our scheme, and to this end we use the simpler scheme $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ in which set elements are prime numbers in $\left(2^{\mu-1}, 2^{\mu}\right)$. The commitment to the set $P = \{e_1, \ldots, e_n\}$ is an RSA accumulator [2, 6] that is defined as $\mathsf{Acc} = G^{\prod_{e_i \in P} e_i}$ for a random quadratic residue $G \in \mathsf{QR}_N$. The commitment to a set element e is instead a Pedersen commitment $c_e = g^e h^{r_q}$ in a group $\mathbb{G}_q$ of prime order q, where q is of ν bits and $\mu < \nu$. For public commitments Acc and c_e, our scheme allows to prove in zero-knowledge the knowledge of e committed in c_e such that $e \in P$ and $\mathsf{Acc} = G^{\prod_{e_i \in P} e_i}$. A public coin protocol for this problem was proposed by Camenisch and Lysyanskaya [16]. Their protocol however requires various restrictions. For instance, the accumulator must work with at least 2λ-bit long primes, which slows down accumulation time, and the prime order group must be more than 4λ-bits (e.g., of 512 bits), which is undesirable for efficiency reasons, especially if this prime order group is used to instantiate more proof systems to create other proofs about the committed element. In our scheme the goal is instead to keep the prime order group of "normal" size (say, 2λ bits), so that it can be for example a prime order group in which we can efficiently instantiate another CP-SNARK that could be composed with our $\mathsf{MemCP}_{\mathsf{RSAPrm}}$. And we can also allow flexible choices of the primes size that can be tuned to the application so that applications that work with moderately large sets can benefit in efficiency. In order to achieve these goals, our idea to create a membership proof is to compute the following:

- An accumulator membership witness $W = G^{\prod_{e_i \in P \setminus \{e\}} e_i}$, and an integer commitment to e in the RSA group, $C_e = G^e H^r$, where $H \in \mathsf{QR}_N$.
- A ZK proof of knowledge $\mathsf{CP}_{\mathsf{Root}}$ of a committed root for Acc, i.e. a proof of knowledge of e and W such that $W^e = \mathsf{Acc}$ and $C_e = G^e H^r$.
  Intuitively, this gives that C_e commits to an integer that is accumulated in Acc (at this point, however, the integer may be a trivial root, i.e., 1).
- A ZK proof $\mathsf{CP}_{\mathsf{modEq}}$ that C_e and c_e commit to the same value modulo q.
- A ZK proof $\mathsf{CP}_{\mathsf{Range}}$ that c_e commits to an integer in the range $\left(2^{\mu-1}, 2^{\mu}\right)$.

From the combination of the above proofs we would like to conclude that the integer committed in c_e is in P. Without further restrictions, however, this may not be the case; in particular, since for the value committed in C_e we do not have a strict bound it may be that the integer committed in c_e is another e_q such $e = e_q \pmod q$ but $e \neq e_q$ over the integers. In fact, the

proof $\mathsf{CP}_{\mathsf{Root}}$ does not guarantee us that C_e commits to a single prime number e, but only that e divides $\prod_{e_i \in P} e_i$, namely e might be a product of a few primes in P or the corresponding negative value, while its residue modulo q may be some value that is not in the set—what we call a "collision". We solve this problem by taking in consideration that e_q is guaranteed by $\mathsf{CP}_{\mathsf{Range}}$ to be in $\left(2^{\mu-1}, 2^{\mu}\right)$ and by enhancing $\mathsf{CP}_{\mathsf{Root}}$ to also prove a bound on e: roughly speaking $|e| < 2^{2\lambda_s+\mu}$ for a statistical security parameter λ_s. Using this information we develop a careful analysis that bounds the probability that such collisions can happen for a malicious e (see Sect. 4.3 for more intuition).

In the following section we formally describe the type-based commitment scheme supported by our CP-SNARK, and a collection of building blocks. Then we present the $\mathsf{MemCP}_{\mathsf{RSA}}$ and $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ CP-SNARKs in Sects. 4.3 and 4.4 respectively, and finally we give instantiations for some of our building blocks in Sect. 4.5.

Remark 4.1 Although we specifically describe our protocols for RSA groups, they generalize to work over any Hidden Order Group with slight modifications. See Appendix 4 for details.

4.2 Preliminaries and building blocks

4.2.1 Notation

Given a set $U = \{u_1, \ldots, u_n\} \subset \mathbb{Z}$ of cardinality n we denote compactly with $\mathsf{prod}_U := \prod_{i=1}^{n} u_i$ the product of all its elements. We use capital letters for elements in an RSA group $\mathbb{Z}_N^*$, e.g., $G, H \in \mathbb{Z}_N^*$. Conversely, we use small letters for elements in a prime order group $\mathbb{G}_q$, e.g., $g, h \in \mathbb{G}_q$. Following this notation, we denote a commitment in a prime order group as $c \in \mathbb{G}_q$, while a commitment in an RSA group as $C \in \mathbb{Z}_N^*$.

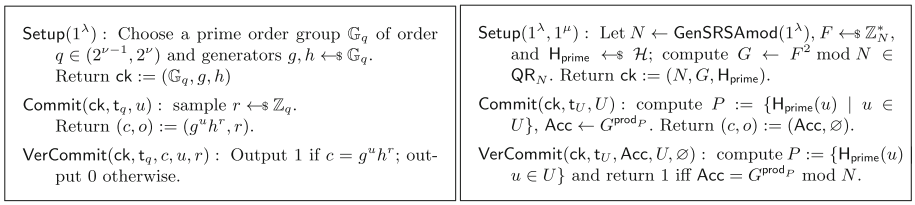
4.2.2 Commitment schemes

Our first CP-SNARK, called $\mathsf{MemCP}_{\mathsf{RSA}}$, is for a family of relations $R_{\mathsf{mem}} : \mathcal{D}_{\mathsf{elm}} \times \mathcal{D}_{\mathsf{set}}$ such that $\mathcal{D}_{\mathsf{elm}} = \{0,1\}^{\eta}$, $\mathcal{D}_{\mathsf{set}} = 2^{\mathcal{D}_{\mathsf{elm}}}$, and for a type-based commitment scheme that is the canonical composition $\mathsf{SetCom}_{\mathsf{RSA}} \bullet \mathsf{PedCom}$ of the two commitment schemes given in Fig. 4. PedCom is essentially a classical Pedersen commitment scheme in a group $\mathbb{G}_q$ of prime order q such that $q \in (2^{\nu-1}, 2^{\nu})$ and $\eta < \nu$. PedCom is used to commit to set elements and its type is t_q. $\mathsf{SetCom}_{\mathsf{RSA}}$ is a (non-hiding) commitment scheme for sets of η-bit strings, that is built as an RSA accumulator [2, 6] to a set of μ-bit primes, each derived from an η-bit string by a deterministic hash function $\mathsf{H}_{\mathsf{prime}} : \{0,1\}^{\eta} \to \mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$. $\mathsf{SetCom}_{\mathsf{RSA}}$ is computationally binding under the factoring assumption[16] and the collision resistance of $\mathsf{H}_{\mathsf{prime}}$. Its type for sets is t_U.

4.2.3 Hashing to primes

The problem of mapping arbitrary values to primes in a collision-resistant manner has been studied in the past, see e.g., [14, 29, 43], and in [40] a method to generate random primes is presented. Although the main idea of our scheme would work with any instantiation of $\mathsf{H}_{\mathsf{prime}}$, for the goal of significantly improving efficiency, our construction considers a specific class

[16] Here is why: finding two different sets of primes $P, P', P \neq P'$ such that $G^{\mathsf{prod}_P} = \mathsf{Acc} = G^{\mathsf{prod}_{P'}}$ implies finding an integer $\alpha = \mathsf{prod}_P - \mathsf{prod}_{P'} \neq 0$ such that $G^{\alpha} = 1$. This is known to lead to an efficient algorithm for factoring N.



(a) PedCom (b) $\mathsf{SetCom}_{\mathsf{RSA}}$

Fig. 4 RSA accumulator and Pedersen commitment schemes for RSAHashmem

of $\mathsf{H}_{\mathsf{prime}}$ functions that work as follows. Let $\mathsf{H} : \{0,1\}^\eta \times \{0,1\}^\iota \to \{0,1\}^{\mu-1}$ be a collision-resistant function, and define $\mathsf{H}_{\mathsf{prime}}(u)$ as the function that starting with $j = 0$, looks for the first $j \in [0, 2^\iota - 1]$ such that the integer represented by the binary string $1|\mathsf{H}(u, j)$ is prime. In case it reaches $j = 2^\iota - 1$ it failed to find a prime and outputs $\perp$.[17] We consider two main candidates of such function H (and thus $\mathsf{H}_{\mathsf{prime}}$):

- *Pseudorandom function* Namely $\mathsf{H}(u, j) := \mathsf{F}_\kappa(u, j)$ where $\mathsf{F}_\kappa : \{0,1\}^{\eta+\iota}$ is a PRF with public seed κ and $\iota = \lceil \log \mu\lambda \rceil$. Due to the density of primes, the corresponding $\mathsf{H}_{\mathsf{prime}}$ runs in the expected running time $O(\mu)$ and $\perp$ is returned with probability $\leq \mathsf{exp}(-\lambda) = \mathsf{negl}(\lambda)$.[18] Under the random oracle heuristic, F can be instantiated with a hash function like SHA256.
- *Deterministic map* $\mathsf{H}(u, j) := f(u) + j$ with $u > 2^{\eta-1}$ and $j \in (f(u), f(u+1))$, where $f(u) := 2(u+2)\log_2(u+1)^2$. The corresponding function $\mathsf{H}_{\mathsf{prime}}(u)$ is essentially the function that maps to the next prime after $f(u)$. This function is collision-free (indeed it requires to take $\mu > \eta$) and generates primes that can be smaller (in expectation) than the function above. Cramer's conjecture implies that the interval $(f(u), f(u+1))$ contains a prime when u is sufficiently large.

4.2.4 CP-NIZK for H computation and PedCom

We use a CP-NIZK $\mathsf{CP}_{\mathsf{HashEq}}$ for the relation $R_{\mathsf{HashEq}} : \{0,1\}^\mu \times \{0,1\}^\eta \times \{0,1\}^\iota$ defined as

$$R_{\mathsf{HashEq}}(u_1, u_2, \omega) = 1 \iff u_1 = (1|\mathsf{H}(u_2, \omega)),$$

and for the commitment scheme PedCom. Essentially, with this scheme one can prove that two commitments c_e and c_u in $\mathbb{G}_q$ are such that $c_e = g^e h^{r_q}$, $c_u = g^u h^{r_u}$ and there exists j such that $e = (1|\mathsf{H}(u, j))$. As it shall become clear in our security proof, we do not have to prove all the iterations of H until finding j such that $(1|\mathsf{H}(u, j)) = \mathsf{H}_{\mathsf{prime}}(u)$ is prime, which saves significantly on the complexity of this CP-NIZK.

4.2.5 Integer commitments

We use a scheme for committing to arbitrarily large integer values in RSA groups introduced by Fujisaki and Okamoto [41] and later improved in [31]. We briefly recall the commitment scheme. Let $\mathbb{Z}_N^*$ be an RSA group. The commitment key consists of two randomly chosen generators $G, H \in \mathbb{Z}_N^*$; to commit to any $x \in \mathbb{Z}$ one chooses randomly an $r \leftarrow\$ [1, N/2]$ and

[17] For specific instantiations of H, ι can be set so that $\perp$ is returned with negligible probability.

[18] We assume for simplicity that the function never outputs $\perp$, though it can happen with negligible probability.

computes $C \leftarrow G^x H^r$; the verifier checks whether or not $C = \pm G^x H^r$. This commitment scheme is statistically hiding, as long as G and H lie in the subgroup of $\mathbb{Z}_N^*$. This can be achieved by setting $G \leftarrow F^2$, $H \leftarrow J^2 \in \mathsf{QR}(N)$, where F, J are randomly sampled from $\mathbb{Z}_N^*$. Moreover it's computationally binding under the assumption that factoring is hard in $\mathbb{Z}_N^*$. Furthermore, a proof of knowledge of an opening was presented in [31], its knowledge soundness was based on the strong RSA assumption, and later found to be reducible to the plain RSA assumption in [25]. We denote this commitment scheme as IntCom.

4.2.6 Strong-RSA accumulators

As observed earlier, our commitment scheme for sets is an RSA accumulator Acc computed on the set of primes P derived from U through the map to primes, i.e., $P := \{\mathsf{H}_{\mathsf{prime}}(s) | s \in U\}$. In our construction we use the accumulator's feature for computing succinct membership witnesses, which we recall works as follows. Given $\mathsf{Acc} = G^{\prod_{e_i \in P} e_i} := G^{\mathsf{prod}_P}$, the membership witness for e_k is $W_k = G^{\prod_{e_i \in P \setminus \{e_k\}} e_i}$, which can be verified by checking if $W_k^{e_k} = \mathsf{Acc}$.

4.2.7 Argument of knowledge of a root

We make use of a zero-knowledge non-interactive argument of knowledge of a root of a public RSA group element $\mathsf{Acc} \in \mathsf{QR}_N$. This NIZK argument is called $\mathsf{CP}_{\mathsf{Root}}$. More precisely, it takes in an integer commitment to a $e \in \mathbb{Z}$ and then proves knowledge of an e-th root of Acc, i.e., of $W = \mathsf{Acc}^{\frac{1}{e}}$. More formally, $\mathsf{CP}_{\mathsf{Root}}$ is a NIZK for the relation $R_{\mathsf{Root}} : (\mathbb{Z}_N^* \times \mathsf{QR}_N \times \mathbb{N}) \times (\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}_N^*)$ defined as
$R_{\mathsf{Root}}((C_e, \mathsf{Acc}, \mu), (e, r, W)) = 1$ iff,

$$C_e = \pm G^e H^r \bmod N \ \wedge \ W^e = \mathsf{Acc} \bmod N \ \wedge \ |e| < 2^{\lambda_z + \lambda_s + \mu + 2},$$

where λ_z and λ_s are the statistical zero-knowledge and soundness security parameters respectively of the protocol $\mathsf{CP}_{\mathsf{Root}}$. $\mathsf{CP}_{\mathsf{Root}}$ is obtained by applying the Fiat–Shamir transform to a public-coin protocol that we propose based on ideas from the protocol of Camenisch and Lysyanskaya for proving knowledge of an accumulated value [16]. In [16], the protocol ensures that the committed integer e is in a specific range, different from 1 and positive. In our $\mathsf{CP}_{\mathsf{Root}}$ protocol we instead removed these constraints and isolated the portion of the protocol that only proves knowledge of a root. We present the $\mathsf{CP}_{\mathsf{Root}}$ protocol in Sect. 4.5; its interactive public coin version is knowledge sound under the RSA assumption and statistical zero-knowledge. Finally, we notice that the relation R_{Root} is defined for statements where $\mathsf{Acc} \in \mathsf{QR}_N$, which may not be efficiently checkable given only N if Acc is adversarially chosen. Nevertheless $\mathsf{CP}_{\mathsf{Root}}$ can be used in larger cryptographic constructions that guarantee $\mathsf{Acc} \in \mathsf{QR}_N$ through some extra information, as is the case in our scheme.

4.2.8 Proof of equality of commitments in $\mathbb{Z}_N^*$ and $\mathbb{G}_q$

Our last building block, called $\mathsf{CP}_{\mathsf{modEq}}$, proves in zero-knowledge that two commitments, a Pedersen commitment in a prime order group and an integer commitment in an RSA group, open to the same value modulo the prime order $q = \mathsf{ord}(\mathbb{G})$. This is a conjunction of a classic Pedersen Σ-protocol and a proof of knowledge of opening of an integer commitment [31], i.e. for the relation

$\mathsf{KeyGen}(\mathsf{ck}, R^{\in}):$ parse $\mathsf{ck} := ((N, G, \mathsf{H}_{\mathsf{prime}}), (\mathbb{G}_q, g, h))$ as the commitment keys of $\mathsf{SetCom}_{\mathsf{RSA}}$ and PedCom respectively. Sample a random generator H.
Generate $\mathsf{crs}_{\mathsf{HashEq}} \leftarrow\$ \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{KeyGen}((\mathbb{G}_q, g, h), R_{\mathsf{HashEq}})$, a crs for $\mathsf{CP}_{\mathsf{HashEq}}$.
Return $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{HashEq}})$.
Given crs, one can define $\mathsf{crs}_{\mathsf{Root}} := (N, G, H)$, $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$.

$\mathsf{Prove}(\mathsf{crs}, (C_U, c_u), (U, u), (\varnothing, r_u)):$ $e \leftarrow \mathsf{H}_{\mathsf{prime}}(u) = (1|\mathsf{H}(u, j))$, $(c_e, r_q) \leftarrow \mathsf{Com}_1.\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_q, e)$.
$(C_e, r) \leftarrow \mathsf{IntCom}.\mathsf{Commit}((G, H), e)$; $P \leftarrow \{\mathsf{H}_{\mathsf{prime}}(u) : u \in U\}$, $W = G^{\prod_{e_i \in P \setminus \{e\}} e_i}$.
$\pi_{\mathsf{Root}} \leftarrow \mathsf{CP}_{\mathsf{Root}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Root}}, (C_e, C_U, \mu), (e, r, W))$
$\pi_{\mathsf{modEq}} \leftarrow \mathsf{CP}_{\mathsf{modEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), (e, e, r, r_q))$
$\pi_{\mathsf{HashEq}} \leftarrow \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{HashEq}}, (c_e, c_u), (e, u), (r_q, r_u), j)$
Return $\pi := (C_e, c_e, \pi_{\mathsf{Root}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{HashEq}})$.

$\mathsf{VerProof}(\mathsf{crs}, (C_U, c_u), \pi):$ Return 1 iff $\mathsf{CP}_{\mathsf{Root}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Root}}, (C_e, C_U, \mu), \pi_{\mathsf{Root}}) = 1 \wedge$
$\mathsf{CP}_{\mathsf{modEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), \pi_{\mathsf{modEq}}) = 1 \wedge \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{HashEq}}, (c_e, c_u), \pi_{\mathsf{HashEq}}) = 1.$

Fig. 5 $\mathsf{MemCP}_{\mathsf{RSA}}$ CP-SNARK for set membership

$$R_{\mathsf{modEq}}\left((C_e, c_e), (e, e_q, r, r_q)\right) = 1 \text{ iff } e = e_q \bmod q \wedge C_e = \pm G^e H^r \bmod N \wedge c_e = g^{e_q \bmod q} h^{r_q \bmod q}.$$

We present $\mathsf{CP}_{\mathsf{modEq}}$ in Sect. 4.5.

4.3 Our CP-SNARK $\mathsf{MemCP}_{\mathsf{RSA}}$

We are now ready to present our CP-SNARK $\mathsf{MemCP}_{\mathsf{RSA}}$ for set membership. The scheme is fully described in Fig. 5 and makes use of the building blocks presented in the previous section.

The KeyGen algorithm takes as input the commitment key of Com_1 and a description of R_{mem} and does the following: it samples a random generator $H \leftarrow\$ \mathsf{QR}_N$ so that (G, H) define a key for the integer commitment, and generate a CRS $\mathsf{crs}_{\mathsf{HashEq}}$ of the $\mathsf{CP}_{\mathsf{HashEq}}$ CP-NIZK.

For generating a proof, the ideas are similar to the ones informally described at the beginning of Sect. 4 for the case when set elements are prime numbers. In order to support sets U of arbitrary strings the main differences are the following: (i) we use $\mathsf{H}_{\mathsf{prime}}$ in order to derive a set of primes P from U, (ii) given a commitment c_u to an element $u \in \{0, 1\}^\eta$, we commit to $e = \mathsf{H}_{\mathsf{prime}}(u)$ in c_e; (iii) we use the previously mentioned ideas to prove that c_e commits to an element in P (that is correctly accumulated), except that we replace the range proof π_{Range} with a proof π_{HashEq} that c_u and c_e commits to u and e respectively, such that $\exists j : e = (1|\mathsf{H}(u, j))$.

Remark 4.2 (On the support of larger η) In order to commit to a set element $u \in \{0, 1\}^\eta$ with the PedCom scheme we require $\eta < \nu$. This condition is actually used for ease of presentation. It is straightforward to extend our construction to the case $\eta \geq \nu$, in which case every u should be split in blocks of less than ν bits that can be committed using the vector Pedersen commitment (Fig. 4).

The correctness of $\mathsf{MemCP}_{\mathsf{RSA}}$ can be checked by inspection: essentially, it follows from the correctness of all the building blocks and the condition that $\eta, \mu < \nu$. For succinctness, we observe that the commitments C_U, c_u and all the three proofs have size that does not depend on the cardinality of the set U, which is the only portion of the witness whose size is not a-priori fixed.

4.3.1 Proof of security

Recall that the goal is to prove in ZK that c_u is a commitment to an element $u \in \{0,1\}^\eta$ that is in a set U committed in C_U. Intuitively, we obtain the security of our scheme from the conjunction of proofs for relations R_{Root}, R_{modEq} and R_{HashEq}: (i) π_{HashEq} gives us that c_e commits to $e_q = (1|\mathsf{H}(u, j))$ for some j and for u committed in c_u. (ii) π_{modEq} gives that C_e commits to an integer e such that $e \mod q = e_q$ is committed in c_e. (iii) π_{Root} gives us that the integer e committed in C_e divides prod_P, where $C_U = G^{\mathsf{prod}_P}$ with $P = \{\mathsf{H}_{\mathsf{prime}}(u_i) : u_i \in U\}$.

By combining these three facts we would like to conclude that $e_q \in P$ that, together with π_{HashEq}, should also guarantee $u \in U$. A first problem to analyze, however, is that for e we do not have guarantees of a strict bound in $\left(2^{\mu-1}, 2^\mu\right)$; so it may in principle occur that $e = e_q \pmod q$ but $e \neq e_q$ over the integers. Indeed, the relation R_{Root} does not guarantee us that e is a single prime number, but only that e divides the product of primes accumulated in C_U. Assuming the hardness of Strong RSA we may still have that e is the product of a few primes in P or even is a negative integer. We expose a simple attack that could arise from this: an adversary can find a product of primes from the set P, let it call e, such that $e = e_q \pmod q$ but $e \neq e_q$ over the integers. Since e is a legitimate product of members of P, the adversary can efficiently compute the e-th root of C_U and provide a valid π_{Root} proof. This is what we informally call a "collision". Another simple attack would be that an adversary takes a single prime e and then commits to its opposite $e_q \leftarrow -e \mod q$ in the prime order group. Again, since $e \in P$ the adversary can efficiently compute the e-th root of C_U, $W^e = C_U$, and then the corresponding $-e$-th root of C_U, $\left(W^{-1}\right)^{-e} = C_U$. This is a second type of attack to achieve what we called "collision". With a careful analysis we show that with appropriate parameters the probability that such collisions occur can be either 0 or negligible.

One key observation is that R_{Root} does guarantee a lower and an upper bound, $-2^{\lambda_z+\lambda_s+\mu+2}$ and $2^{\lambda_z+\lambda_s+\mu+2}$ respectively, for e committed in C_e. From these bounds (and that $e \mid \mathsf{prod}_P$) we get that an adversarial e can be the product of *at most* $d = 1 + \lfloor \frac{\lambda_z+\lambda_s+2}{\mu} \rfloor$ primes in P (or their corresponding negative product). Then, if $2^{d\mu} \leq 2^{\nu-2} < q$, or $d\mu + 2 \leq \nu$, we get that $e < 2^{d\mu} < q$. In case $e > 0$ and since q is prime, $e = e_q \bmod q \wedge e < q$ implies that $e = e_q$ over $\mathbb{Z}$, namely no collision can occur at all. In the other case $e < 0$ we have $e > -2^{d\mu}$ and $e = e_q \pmod q$ implies $e = -q + e_q < -2^{\nu-1} + 2^\mu < -2^{\nu-1} + 2^{\nu-2} = -2^{\nu-2}$. Therefore, $-2^{d\mu} < -2^{\nu-2}$, which is a contradiction since we assumed $d\mu + 2 \leq \nu$. So this type of collision cannot happen.

If on the other hand we are in a parameters setting where $d\mu > \nu - 2$, we give a concrete bound on the probability that such collisions occur. More precisely, for this case we need to assume that the integers returned by H are random, i.e., H is a random oracle, and we also use the implicit fact that R_{HashEq} guarantees that $e_q \in \left(2^{\mu-1}, 2^\mu\right)$. Then we give a concrete bound on the probability that the product of d out of $\mathsf{poly}(\lambda)$ random primes lies in a specific range $\left(2^{\mu-1}, 2^\mu\right)$, which turns out to be negligible when d is constant and $2^{\mu-\nu}$ is negligible.

Since the requirements of security are slightly different according to the setting of parameters mentioned above, we state two separate theorems, one for each case.

Theorem 4.1 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}}$ *and* IntCom *be computationally binding commitments,* CP_{Root}, CP_{modEq} *and* CP_{HashEq} *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption holds, and that* H *is collision resistant. If* $d\mu + 2 \leq \nu$, *then* $\mathsf{Mem}CP_{\mathsf{RSA}}$ *is knowledge-sound with partial opening of the set commitments* C_U.

Theorem 4.2 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}}$ *and* IntCom *be computationally binding commitments,* $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{HashEq}}$ *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption hold, and that* H *is collision resistant. If* $d\mu + 2 > \nu$, $d = O(1)$ *is a small constant,* $2^{\mu-\nu} \in \mathsf{negl}(\lambda)$ *and* H *is modeled as a random oracle, then* $\mathsf{MemCP}_{\mathsf{RSA}}$ *is knowledge-sound with partial opening of the set commitments* C_U.

Remark 4.3 It is worth noting that Theorem 4.2 where we assume H to be a random oracle requires a random oracle assumption stronger than usual; this has to do with the fact that while we assume H to be a random oracle we also assume that $\mathsf{CP}_{\mathsf{modEq}}$ can create proof about correct computations of H. Similar assumptions have been considered in previous works, see, e.g, [71, Remark 2].

Finally, we state the theorem about the zero-knowledge of $\mathsf{MemCP}_{\mathsf{RSA}}$.

Theorem 4.3 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}}$ *and* IntCom *be statistically hiding commitments,* $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{HashEq}}$ *be zero-knowledge arguments. Then* $\mathsf{MemCP}_{\mathsf{RSA}}$ *is zero-knowledge.*

Proof (Sketch) The proof is rather straightforward, so we only provide a sketch. We define the simulator $\mathcal{S}$ that takes as input (crs, C_U, c_u) and does the following:

- Parses $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{HashEq}})$, from which it computes the corresponding $\mathsf{crs}_{\mathsf{Root}} := (N, G, H)$ and $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$.
- Samples at random $C_e^* \leftarrow_\$ \mathbb{Z}_N^*$ and $c_e^* \leftarrow_\$ \mathbb{G}_q$.
- Invokes $\mathcal{S}_{\mathsf{Root}}(\mathsf{crs}_{\mathsf{Root}}, C_e^*, C_U)$, $\mathcal{S}_{\mathsf{modEq}}(\mathsf{crs}_{\mathsf{modEq}}, C_e^*, c_e^*)$ and $\mathcal{S}_{\mathsf{HashEq}}(\mathsf{crs}_{\mathsf{HashEq}}, c_e^*, c_u)$ the corresponding simulators of $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{HashEq}}$ respectively. They output simulated proof π^*_{Root}, π^*_{modEq} and π^*_{HashEq} respectively.
- $\mathcal{S}$ outputs $(C_e^*, c_e^*, \pi^*_{\mathsf{Root}}, \pi^*_{\mathsf{modEq}}, \pi^*_{\mathsf{HashEq}})$.

Let $\pi := (C_e, c_e, \pi_{\mathsf{Root}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{HashEq}}) \leftarrow \mathsf{Prove}(\mathsf{crs}, (C_U, c_u), (U, u), (\varnothing, r_u))$ be the output of a real proof. Since IntCom and PedCom are statistically hiding C_e^* and c_e^* are indistinguishable from C_e and c_e resp. Finally, since $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{HashEq}}$ are zero knowledge arguments π^*_{Root}, π^*_{modEq} and π^*_{HashEq} are indistinguishable from π_{Root}, π_{modEq} and π_{HashEq} resp. □

4.3.2 Notation

We introduce some notation that eases our proofs exposition. Let $U = \{u_1, \ldots, u_n\} \subset \mathbb{Z}$ be a set of cardinality n. We denote as prod a product of (an arbitrary number of) elements of U, $\mathsf{prod} = \prod_{i \in I} u_i$, for some $I \subseteq [n]$. Furthermore, $\Pi_U = \{\mathsf{prod}_1, \ldots, \mathsf{prod}_{2^n-1}\}$ is the set of all possible products and more specifically $\Pi_{U,d} \subseteq \Pi_U$ denotes the set of possible products of exactly d elements of U, $|I| = d$, while for the degenerate case of $d > n$ we define $\Pi_{U,d} = \emptyset$. We note that $|\Pi_{U,d}| = \binom{n}{d}$ (except for the degenerate case where $|\Pi_{U,d}| = 0$). For convenience, in the special case of $\mathsf{prod} \in \Pi_{U,|U|}$, i.e. the (unique) product of all elements of U, we will simply write prod_U. Finally, for a $J \subseteq [n]$ we let $\Pi_{U,J} = \cup_{j \in J} \Pi_{U,j}$; for example $\Pi_{U,[1,\ldots,d]} = \cup_{j=1}^{d} \Pi_{U,j}$ is the set of all possible products of up to d elements of U. For all of the above we also denote with "$-$" the corresponding set of the opposite element, e.g. $-\Pi_U = \{-\mathsf{prod}_1, \ldots, -\mathsf{prod}_{2^n-1}\}$

Proof of Theorem 4.1 Let a malicious prover $\mathcal{P}^*$, a PPT adversary of Knowledge Soundness with Partial Opening (see the definition in Sect. 2.6) that on input $(\mathsf{ck}, R_{\mathsf{mem}}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$

outputs (C_U, c_u, U, π) such that the verifier $\mathcal{V}$ accepts, i.e. $\mathsf{VerProof}(\mathsf{crs}, C_U, c_u), \pi) = 1$ and $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, C_U, U, \varnothing) = 1$ with non-negligible probability ϵ. We will construct a PPT extractor $\mathcal{E}$ that on the same input outputs a partial witness (u, r_q) such that $R_{\mathsf{mem}}(U, u) = 1 \wedge \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c_u, u, r_q) = 1$.

For this we rely on the Knowledge Soundness of $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{HashEq}}$ protocols. $\mathcal{E}$ parses $\pi := (C_e, c_e, \pi_{\mathsf{Root}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{HashEq}})$ and $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{HashEq}})$, from which it computes the corresponding $\mathsf{crs}_{\mathsf{Root}} := (N, G, H)$ and $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$. Then constructs an adversary $\mathcal{A}_{\mathsf{Root}}$ for $\mathsf{CP}_{\mathsf{Root}}$ Knowledge Soundness that outputs $(C_e, C_U, \mu, \pi_{\mathsf{Root}})$. It is obvious that since $\mathcal{V}$ accepts π then it also accepts π_{Root}, i.e., $\mathsf{CP}_{\mathsf{Root}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Root}}, (C_e, C_U, \mu), \pi_{\mathsf{Root}}) = 1$. From Knowledge Soundness of $\mathsf{CP}_{\mathsf{Root}}$ we know that there is an extractor $\mathcal{E}_{\mathsf{Root}}$ that outputs (e, r, W) such that $C_e = \pm G^e H^r \pmod N \wedge W^e = C_U \pmod N \wedge |e| < 2^{\lambda_z+\lambda_s+\mu+2}$. Similarly, $\mathcal{E}$ constructs adversaries $\mathcal{A}_{\mathsf{modEq}}$ and $\mathcal{A}_{\mathsf{HashEq}}$ of protocols $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{HashEq}}$ respectively. And similarly there are extractors $\mathcal{E}_{\mathsf{modEq}}$ and $\mathcal{E}_{\mathsf{HashEq}}$ that output (e', e_q, r', r_q) such that $e' = e_q \pmod q \wedge C_{e'} = \pm G^{e'} H^{r'} \pmod N \wedge c_{e_q} = g^{e_q \bmod q} h^{r_q \bmod q}$ and (e'_q, u, r'_q, r_u, j) such that $c_e = g^{e'_q} h^{r'_q} \wedge e'_q = (1|\mathsf{H}(u, j))$ respectively.

From the Binding property of the integer commitment scheme we get that $e = e'$ and $r = r'$ (over the integers), unless with a negligible probability. Similarly, from the Binding property of the Pedersen commitment scheme we get that $e_q = e'_q \pmod q$ and $r_q = r'_q \pmod q$, unless with a negligible probability. So if we put everything together the extracted values are $(e, r, W, e_q, r_q, u, r_u, j)$ such that:

$$W^e = C_U \pmod N \wedge |e| < 2^{\lambda_z+\lambda_s+\mu+2} \wedge e = e_q \pmod q \wedge e_q = (1|\mathsf{H}(u, j)),$$

and additionally

$$C_e = \pm G^e H^r \wedge c_e = g^{e_q \bmod q} h^{r_q \bmod q} \wedge \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c_u, u, r_u) = 1.$$

From $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, C_U, U, \varnothing) = 1$ we infer that $C_U = G^{\mathsf{prod}_P}$, where $P := \{\mathsf{H}_{\mathsf{prime}}(u) \mid u \in U\}$. From the strong RSA assumption since $W^e = C_U = G^{\mathsf{prod}_P} \pmod N$ we get $e \in \Pi_P$ or $e \in -\Pi_P$, unless with a negligible probability (see Appendix 2).

Since, all the elements of P are outputs of $\mathsf{H}_{\mathsf{prime}}$ they have exactly bitlength μ, that is $2^{\mu-1} < e_i < 2^\mu$ for each $e_i \in P$. This means that e is a $(\pm)$ product of μ-sized primes. Let $|e|$ be a product of ℓ primes, meaning that $2^{\ell(\mu-1)} < |e| < 2^{\ell\mu}$, and $d := \lfloor \frac{\lambda_z+\lambda_s+\mu+2}{\mu} \rfloor$. From $|e| < 2^{\lambda_z+\lambda_s+\mu+2}$ we get that $2^{\ell\mu} < 2^{\lambda_z+\lambda_s+\mu+2} \Rightarrow \ell < d$ which means that $e \in \Pi_{P,[1,\ldots,d]}$ or $e \in -\Pi_{P,[1,\ldots,d]}$ (i.e. e is a $(\pm)$ product of at most d primes).

First we show that $e \in \Pi_P$, i.e., that e cannot be negative. Let $e \in -\Pi_{P,[1,\ldots,d]}$. We use the fact that $e = e_q \pmod q$, so $e \leq -q + e_q < -2^{\nu-1} + 2^\mu < -2^{\nu-1} + 2^{\nu-2} = -2^{\nu-2}$. Since $-2^{d\mu} < e$ this leads to $-2^{d\mu} < -2^{\nu-2}$ which contradicts the assumption $d\mu + 2 \leq \nu$ (we used the fact that $e_q = (1|\mathsf{H}(u, j))$ to conclude that $2^{\mu-1} < e_q < 2^\mu$, which comes from the definition of H). So $e > 0$ or $e \in \Pi_{P,[1,\ldots,d]}$.

Recall that $e < 2^{d\mu}$. From the assumption $d\mu + 2 \leq \nu$ which means that $e < 2^{d\mu} < 2^{\nu-2} < q \Rightarrow e < q$. Since $e = e_q \pmod q$ and $e < q$ this means that $e = e_q$ over the integers. Again we are using the fact that $e_q = (1|\mathsf{H}(u, j))$ to conclude that $2^{\mu-1} < e_q < 2^\mu$, which comes from the definition of H, and combined with $e = e_q$ we get that $2^{\mu-1} < e < 2^\mu$. The last fact means that $e \in \Pi_{P,\{1\}}$ (i.e. e is exactly one prime from P) otherwise it would exceed 2^μ, so $e \in P$.

Finally, $e = e_q = (1|\mathsf{H}(u, j)) = \mathsf{H}_{\mathsf{prime}}(u) \in P = \{\mathsf{H}_{\mathsf{prime}}(u_1), \ldots, \mathsf{H}_{\mathsf{prime}}(u_n)\}$, where $U := \{u_1, \ldots, u_n\}$. This means that there is an i such that $\mathsf{H}_{\mathsf{prime}}(u) = \mathsf{H}_{\mathsf{prime}}(u_i)$. From colli-

sion resistance of $\mathsf{H}_{\mathsf{prime}}$ we infer that $u = u_i$. So we conclude that $u \in U$ or $R_{\mathsf{mem}}(U, u) = 1$ and as shown above $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c_u, u, r_u) = 1$. □

4.3.3 Collision finding analysis

For the second theorem we cannot count on the formula $d\mu + 2 \leq \nu$ that ensures that the extracted integer e lies inside $[0, q-1]$. As explained above, we can only rely on the randomness of each prime to avoid the described "collisions". First, we formally define what a "collision" is through a probabilistic experiment, CollisionFinding, and then we compute a concrete bound for the probability that this event happens, i.e. the experiment outputs 1. Finally, we state a theorem that shows this probability is asymptotically negligible under the assumption that $2^{\mu-\nu}$ is a negligible value (and d is a constant).

$\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, n)$

Let $P \subseteq \mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$ be a randomly chosen set of cardinality n, i.e. each $e \in P$ is chosen uniformly at random, $e_i \leftarrow\!\!\$\ \mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$ meaning that:

1. e_i is prime.
2. $2^{\mu-1} \leq e_i \leq 2^{\mu}$
3. $Pr[e_i = x] = \frac{\mu}{2^{\mu}} + \mathsf{negl}(\lambda)$ for each $x \in \mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$

The output of the experiment is 1 iff there exists $\mathsf{prod} \in \left(\Pi_{P,[2,d]} \cup -\Pi_{P,[2,d]}\right)$ such that $(\mathsf{prod} \mod q) \in \left(2^{\mu-1}, 2^{\mu}\right)$

Lemma 4.1 *Let $\mathbb{G}_q$ be a prime order group of order $q \in \left(2^{\nu-1}, 2^{\nu}\right)$ and μ such that $\mu < \nu$ then* $Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_e, n) = 1] \leq 2 \cdot \sum_{j=2}^{d} \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}$.

Proof First we will prove it for positive products, that is we bound the probability $Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_e, n) = 1 | \mathsf{prod} \in \Pi_{P,[2,d]}]$. Let $\mathsf{prod} = q_1 ... q_j$ be a product of exactly j primes for a $2 \leq j \leq d$. Since $q_i \in \left(2^{\mu-1}, 2^{\mu}\right)$ we get $\mathsf{prod} = q_1 ... q_j \in \left(2^{j\mu-j}, 2^{j\mu}\right)$. Also $\mathbb{Z}_q^*$ is cyclic so we know that at most

$$\left\lceil \frac{\left|\left(2^{j\mu-j}, 2^{j\mu}\right)\right|}{q} \right\rceil = \left\lceil \frac{2^{j\mu} - 2^{j\mu-j}}{q} \right\rceil = \left\lceil \frac{2^{j\mu-j} \cdot (2^j - 1)}{q} \right\rceil \leq 2^{j\mu-j-\nu+1} \cdot (2^j - 1),$$

integers in $\left(2^{j\mu-j}, 2^{j\mu}\right)$ are equal to c modulo q, for any $c \in \{0, 1, ..., q-1\}$.

We are interested in the interval $\left(2^{\mu-1}, 2^{\mu}\right)$ modulo q. From the previous we get that at most $2^{j\mu-j-\nu+1} \cdot (2^j - 1) \cdot \left|\left(2^{\mu-1}, 2^{\mu}\right)\right| = 2^{j\mu-j-\nu+1} \cdot (2^j - 1) \cdot 2^{\mu-1} = 2^{(j+1)\mu-j-\nu}(2^j - 1)$ integers in the range of $\left(2^{j\mu-j}, 2^{j\mu}\right)$ are "winning" integers for the adversary, meaning that after modulo q they are mapped to the winning interval $\left(2^{\mu-1}, 2^{\mu}\right)$.

From the distribution of primes we know that the number of primes in $\left(2^{\mu-1}, 2^{\mu}\right)$ is approximately $\frac{2^{\mu-1}}{\mu-1}$. So there are (approximately) $\left(\frac{2^{\mu-1}}{\mu-1}\right)^j = \frac{2^{j\mu-j}}{(\mu-1)^j}$ different products of j primes from $\mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$ in $\left(2^{j\mu-d}, 2^{j\mu}\right)$.

This leads us to the combinatorial experiment of choice of $B = \frac{2^{j\mu-j}}{(\mu-1)^j}$ "balls", with $T = 2^{(j+1)\mu-j-\nu}(2^j - 1)$ "targets" and $X = \binom{n}{j}$ "tries" without replacement, where "balls"

are all possible products, "targets" are the ones that go to $\left(2^{\mu-1}, 2^{\mu}\right)$ modulo q (the winning ones) and tries are the number of products (for a constant j) that the adversary can try. The "without replacement" comes from the fact that all products are different. The final winning probability is:

$$\begin{aligned} Pr[\mathsf{prod} \mod q \in \left(2^{\mu-1}, 2^{\mu}\right) \wedge \mathsf{prod} \in \Pi_{P,j}] &\leq \frac{T}{B} + \frac{T}{B-1} + \frac{T}{B-2} + \ldots + \frac{T}{B-X} \\ &\leq X \cdot \frac{T}{B-X} \\ &= \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}. \end{aligned}$$

By applying the union bound for all j's we get:

$$Pr[\mathsf{prod} \mod q \in \left(2^{\mu-1}, 2^{\mu}\right) \wedge \mathsf{prod} \in \Pi_{P,[2,d]}] \leq \sum_{j=2}^{d} \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}.$$

By using the same arguments for negative products we would conclude that

$$Pr[\mathsf{prod} \mod q \in \left(2^{\mu-1}, 2^{\mu}\right) \wedge \mathsf{prod} \in -\Pi_{P,[2,d]}] \leq \sum_{j=2}^{d} \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}.$$

Therefore

$$\begin{aligned} &Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_e, n) = 1] \\ &= Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_e, n) = 1 \wedge \mathsf{prod} \in \Pi_{P,[2,d]}] + \\ &\quad + Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_e, n) = 1 \wedge \mathsf{prod} \in -\Pi_{P,[2,d]}] = \\ &\leq 2 \cdot \sum_{j=2}^{d} \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}. \end{aligned}$$

□

Theorem 4.4 *Let $\mathbb{G}_q$ be a prime order group of order $q \in \left(2^{\nu-1}, 2^{\nu}\right)$, μ such that $2^{\mu-\nu} \in \mathsf{negl}(\lambda)$, d constant and $n = \mathsf{poly}(\lambda)$ then $Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, n) = 1] \in \mathsf{negl}(\lambda)$*

Proof Now $n = \mathsf{poly}(\lambda)$ so the set P is polynomially bounded. Due to Lemma 4.1 it is straightforward that $Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, n) = 1] \leq \sum_{j=2}^{d} \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}}$. Since d is constant, for any $j \in [2, d]$ $\binom{n}{j} = O(n^j)$ and we get:

$$\begin{aligned} 2 \cdot \frac{\binom{n}{j} 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - \binom{n}{j}} &= 2 \cdot \frac{O(n^j) 2^{(j+1)\mu-j-\nu}(2^j-1)}{\frac{2^{j\mu-j}}{(\mu-1)^j} - O(n^j)} \\ &= 2 \cdot \frac{O(n^j)(2^j-1)(\mu-1)^j}{\frac{2^{j\mu-j}}{2^{(j+1)\mu-j-\nu}} - \frac{O(n^j)(\mu-1)^j}{2^{(j+1)\mu-j-\nu}}}. \end{aligned}$$

$O(n^j)(2^j-1)(\mu-1)^j = \mathsf{poly}(\lambda)$ and $\frac{O(n^j)(\mu-1)^j}{2^{(j+1)\mu-j-\nu}} = \mathsf{negl}(\lambda)$. Also $\frac{2^{j\mu-j}}{2^{(j+1)\mu-j-\nu}} = 2^{\nu-\mu}$, therefore for j we get a probability bounded by $\frac{\mathsf{poly}(\lambda)2^{\mu-\nu}}{1-\mathsf{negl}(\lambda)2^{\mu-\nu}} = \mathsf{negl}(\lambda)$ by assumption.

$\mathsf{Setup}(1^\lambda, 1^\mu)$: Sample an RSA modulus $N \leftarrow \mathsf{GenSRSAmod}(1^\lambda)$, a random group element $F \leftarrow\!\!\$\ \mathbb{Z}_N^*$, compute $G \leftarrow F^2 \bmod N \in \mathsf{QR}_N$. Return $\mathsf{ck} := (N, G)$.

$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_U, U)$: compute $\mathsf{Acc} \leftarrow G^{\mathsf{prod}_P}$. Return $(c, o) := (\mathsf{Acc}, \varnothing)$.

$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, \mathsf{Acc}, U, \varnothing)$: Return 1 if for all $e_i \in P$ $e_i \in \mathsf{Primes}\left(2^{\mu-1}, 2^\mu\right)$ and $\mathsf{Acc} = G^{\mathsf{prod}_P} \bmod N$, and 0 otherwise.

Fig. 6 $\mathsf{SetCom}_{\mathsf{RSA}'}$ commitment to sets

Finally, $Pr[\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, n) = 1] \leq (d-1) \cdot \mathsf{negl}(\lambda) = \mathsf{negl}(\lambda)$. □

Remark 4.4 For the sake of generality, in $\mathsf{CollisionFinding}$ we do not specify how the random primes are generated. In practice in our scheme they are outputs of the hash function $\mathsf{H}_{\mathsf{prime}}$ that we model as a random oracle.

Now we are ready to give the proof of Theorem 4.2:

Proof of Theorem 4.2 The proof is almost the same as the one of Theorem 4.1 except for the next-to-last paragraph, i.e. the justification of $e \in \Pi_{P,\{1\}}$. Since $d\mu + 2 > \nu$ we cannot use the same arguments to conclude to it. However, still $e \in \left(\Pi_{P,[1,\ldots,d]} \cup -\Pi_{P,[1,\ldots,d]}\right)$.

Let $e \in \left(\Pi_{P,[1,\ldots,d]} \cup -\Pi_{P,[1,\ldots,d]}\right)$, it is straightforward to reduce this case to the the collision finding problem. Assume that the adversary $\mathcal{P}^*$ made q_{H} random oracle queries to H and let Q_{H} be the set of answers she received. Further assume that exactly $q_{\mathsf{H}_{\mathsf{prime}}}$ of the them are primes and let $Q_{\mathsf{H}_{\mathsf{prime}}}$ be the set of them. We note that $P \subseteq Q_{\mathsf{H}_{\mathsf{prime}}}$, unless a collision happened in H.

Now let $Q_{\mathsf{H}_{\mathsf{prime}}}$ be the set of the $\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, |Q_{\mathsf{H}_{\mathsf{prime}}}|)$ experiment. It satisfies all three conditions since each $e_i \in Q_{\mathsf{H}_{\mathsf{prime}}}$ is an output of $\mathsf{H}_{\mathsf{prime}}$. Therefore e_i is prime, $2^{\mu-1} < e_i < 2^\mu$ and since H is modeled as a random oracle the outputs of $\mathsf{H}_{\mathsf{prime}}$ are uniformly distributed in $\mathsf{Primes}\left(2^{\mu-1}, 2^\mu\right)$. Then for the extracted e, we know that $e = e_q \pmod q \in \left(2^{\mu-1}, 2^\mu\right)$ and from the assumption $e \in \left(\Pi_{P,[1,\ldots,d]} \cup -\Pi_{P,[1,\ldots,d]}\right)$, which (as noted above) means that $e \in \left(\Pi_{Q_{\mathsf{H}_{\mathsf{prime}}},[2,\ldots,d]} \cup -\Pi_{Q_{\mathsf{H}_{\mathsf{prime}}},[2,\ldots,d]}\right)$. So $\mathsf{CollisionFinding}(\mu, d, \mathbb{G}_q, |Q_{\mathsf{H}_{\mathsf{prime}}}|) = 1$. Since the adversary is PPT $|Q_{\mathsf{H}_{\mathsf{prime}}}| = \mathsf{poly}(\lambda)$. Also, $d = O(1)$ and $2^{\mu-\nu} \in \mathsf{negl}(\lambda)$ (from the assumptions of the theorem) so the previous happens with a negligible probability according to theorem 4.4. So we conclude that, unless with a negligible probability, $e \in \Pi_{P,\{1\}}$. □

4.4 Our CP-SNARK for set membership for primes sets

In this section we show a CP-SNARK for set membership $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ that supports set elements that are prime numbers of exactly μ bits, i.e., $\mathcal{D}_{\mathsf{elm}} = \mathsf{Primes}(2^{\mu-1}, 2^\mu)$, and $\mathcal{D}_{\mathsf{set}} = 2^{\mathcal{D}_{\mathsf{elm}}}$. $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ works for a type-based commitment scheme Com_2 that is the canonical composition $\mathsf{SetCom}_{\mathsf{RSA}'} \bullet \mathsf{PedCom}$ where $\mathsf{SetCom}_{\mathsf{RSA}'}$ is in Fig. 6 (it is essentially a simplification of $\mathsf{SetCom}_{\mathsf{RSA}}$ since elements are already primes).

The scheme $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ is described in Fig. 7. Its building blocks are the same as the ones for $\mathsf{MemCP}_{\mathsf{RSA}}$ except that instead of a CP-NIZK for proving correctness of a map-to-prime computation, we use a CP-NIZK for range proofs. Namely, we let $\mathsf{CP}_{\mathsf{Range}}$ be a NIZK for the following relation on PedCom commitments c and two given integers $A < B$:

$$R_{\mathsf{Range}}\left((c_e, A, B), (e, r_q)\right) = 1 \ \text{ iff } \ c = g^e h^{r_q} \ \wedge \ A < e_q < B$$

$\mathsf{KeyGen}(\mathsf{ck}, R^{\in})$: parse $\mathsf{ck} := ((N,G),(\mathbb{G}_q,g,h))$ as the commitment keys of $\mathsf{SetCom}_{\mathsf{RSA}'}$ and PedCom respectively. Sample a random generator H.
Generate $\mathsf{crs}_{\mathsf{Range}} \leftarrow\!\!\$\ \mathsf{CP}_{\mathsf{Range}}.\mathsf{KeyGen}((\mathbb{G}_q,g,h),R_{\mathsf{Range}})$, a crs for $\mathsf{CP}_{\mathsf{Range}}$.
Return $\mathsf{crs} := (N,G,H,\mathbb{G}_q,g,h,\mathsf{crs}_{\mathsf{Range}})$.
Given crs, one can define $\mathsf{crs}_{\mathsf{Root}} := (N,G,H)$, $\mathsf{crs}_{\mathsf{modEq}} := (N,G,H,\mathbb{G}_q,g,h)$.
$\mathsf{Prove}(\mathsf{crs},(C_P,c_e),(P,e),(\varnothing,r_q))$:
$(C_e,r) \leftarrow \mathsf{IntCom.Commit}((G,H),e)$
$W = G^{\prod_{e_i\in P\setminus\{e\}} e_i}$.
$\pi_{\mathsf{Root}} \leftarrow \mathsf{CP}_{\mathsf{Root}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Root}},(C_e,C_P,\mu),(e,r,W))$
$\pi_{\mathsf{modEq}} \leftarrow \mathsf{CP}_{\mathsf{modEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{modEq}},(C_e,c_e),(e,e,r,r_q))$
$\pi_{\mathsf{Range}} \leftarrow \mathsf{CP}_{\mathsf{Range}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Range}},(2^{\mu-1},2^{\mu}),c_e,e,r_q)$
Return $\pi := (C_e,\pi_{\mathsf{Root}},\pi_{\mathsf{modEq}},\pi_{\mathsf{Range}})$.
$\mathsf{VerProof}(\mathsf{crs},(C_P,c_e),\pi)$: Return 1 iff
$\mathsf{CP}_{\mathsf{Root}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Root}},(C_e,C_P,\mu),\pi_{\mathsf{Root}}) = 1 \wedge \mathsf{CP}_{\mathsf{modEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{modEq}},(C_e,c_e),\pi_{\mathsf{modEq}}) = 1 \wedge$
$\mathsf{CP}_{\mathsf{Range}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Range}},c_e,\pi_{\mathsf{Range}}) = 1$.

Fig. 7 $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ CP-SNARK for set membership

.

The idea behind the security of the scheme is similar to the one of the $\mathsf{MemCP}_{\mathsf{RSA}}$ scheme. The main difference is that here we rely on the range proof π_{Range} in order to "connect" the Pedersen commitment c_e to the accumulator. In particular, in order to argue the absence of possible collisions here we assume that $d\mu + 2 \leq \nu$ holds, namely we argue security only for this setting of parameters. It is worth noting that in applications where $\mathsf{D}_{\mathsf{elm}}$ is randomly chosen subset of $\mathsf{Primes}\left(2^{\mu-1},2^{\mu}\right)$, we could argue security even when $d\mu + 2 > \nu$, in a way similar to Theorem 4.2. We omit the analysis of this case from the paper.

Theorem 4.5 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}'}$ *and* IntCom *be computationally binding commitments,* $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{Range}}$ *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption hold. If* $d\mu+2 \leq \nu$, *then* $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ *is knowledge-sound with partial opening of the set commitments* c_P. *Furthermore, if* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}'}$ *and* IntCom *are statistically hiding commitments, and* $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{Range}}$ *be zero-knowledge, then* $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ *is zero-knowledge.*

Proof of Theorem 4.5 *K*nowledge soundness with partial opening of C_P: the proof is similar to the one of Theorem 4.1 except for some minor parts.

Let a malicious prover $\mathcal{P}^*$, a PPT adversary of Knowledge Soundness with Partial Opening (see the definition in Sect. 2.6) that on input $(\mathsf{ck}, R_{\mathsf{mem}}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$ outputs (C_P, c_e, P, π) such that the verifier $\mathcal{V}$ accepts, i.e. $\mathsf{VerProof}(\mathsf{crs}, C_P, c_e), \pi) = 1$ and $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, C_P, P, \varnothing) = 1$ with non-negligible probability ϵ. We will construct a PPT extractor $\mathcal{E}$ that on the same input outputs a partial witness (e, r) such that $R_{\mathsf{mem}}(P, e) = 1 \wedge \mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c_e, e, r) = 1$.

For this we rely on the Knowledge Soundness of $\mathsf{CP}_{\mathsf{Root}}$, $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{Range}}$ protocols. $\mathcal{E}$ parses $\pi := (C_e, \pi_{\mathsf{Root}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{Range}})$ and $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{Range}})$, from which it computes the corresponding $\mathsf{crs}_{\mathsf{Root}} := (N, G, H)$ and $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$. Then constructs an adversary $\mathcal{A}_{\mathsf{Root}}$ for $\mathsf{CP}_{\mathsf{Root}}$ Knowledge Soundness that outputs $(C_e, C_P, \mu, \pi_{\mathsf{Root}})$. It is obvious that since $\mathcal{V}$ accepts π then it also accepts π_{Root}, i.e., $\mathsf{CP}_{\mathsf{Root}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Root}}, (C_e, C_P, \mu), \pi_{\mathsf{Root}}) = 1$. From Knowledge Soundness of $\mathsf{CP}_{\mathsf{Root}}$ we know that there is an extractor $\mathcal{E}_{\mathsf{Root}}$ that outputs (e, r, W) such that $C_e = \pm G^e H^r \pmod N \wedge W^e = C_P \pmod N \wedge e < 2^{\lambda_z+\lambda_s+\mu+2}$. Similarly, $\mathcal{E}$ constructs adversaries $\mathcal{A}_{\mathsf{modEq}}$ and $\mathcal{A}_{\mathsf{Range}}$ of protocols $\mathsf{CP}_{\mathsf{modEq}}$ and $\mathsf{CP}_{\mathsf{Range}}$ respectively. And similarly there are extractors $\mathcal{E}_{\mathsf{modEq}}$ and $\mathcal{E}_{\mathsf{Range}}$ that output (e', e_q, r', r_q) such that $e' = e_q$

$(\text{mod } q) \wedge C_{e'} = \pm G^{e'} H^{r'} \ (\text{mod } N) \wedge c_{e_q} = g^{e_q \bmod q} h^{r_q \bmod q}$ and (e'_q, r'_q) such that $c_e = g^{e'_q} h^{r'_q} \wedge 2^{\mu-1} < e'_q < 2^{\mu}$ respectively.

From the Binding property of the integer commitment scheme we get that $e = e'$ and $r = r'$ (over the integers), unless with a negligible probability. Similarly, from the Binding property of the Pedersen commitment scheme we get that $e_q = e'_q \ (\text{mod } q)$ and $r_q = r'_q \ (\text{mod } q)$, unless with a negligible probability. So if we put everything together the extracted values are (e, r, W, e_q, r_q) such that:

$$W^e = C_P \pmod N \wedge e < 2^{\lambda_z+\lambda_s+\mu+2} \wedge e = e_q \pmod q \wedge 2^{\mu-1} < e_q < 2^{\mu},$$

and additionally

$$C_e = \pm G^e H^r \wedge c_e = g^{e_q \bmod q} h^{r_q \bmod q}.$$

From $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, C_P, P, \varnothing) = 1$ we infer that $C_P = G^{\mathsf{prod}_P}$, where for each $e_i \in P$ it holds that $e \in \mathsf{Primes}\left(2^{\mu-1}, 2^{\mu}\right)$. From the strong RSA assumption since $W^e = C_P = G^{\mathsf{prod}_P} \ (\text{mod } N)$ we get $e \in \Pi_P$, unless with a negligible probability (see Appendix 2).

The rest of the analysis that justifies $e \in P$ is identical to the one of the proof of Theorem 4.1. So $e \in P$ and as shown above $\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c_e, e_q, r_q) = 1$.

Zero knowledge For the Zero Knowledge Property we rely on similar techniques with the ones of the proof of Theorem 4.3 except for the use of $\mathcal{S}_{\mathsf{HashEq}}$. Here we use instead the simulator of the $\mathsf{CP}_{\mathsf{Range}}$ protocol, $\mathcal{S}_{\mathsf{Range}}$. □

4.5 Proposed instantiations of protocols for R_{Root} and R_{modEq}

4.5.1 Protocol $\mathsf{CP}_{\mathsf{Root}}$

We first give a protocol $\mathsf{CP}_{\mathsf{Root}'}$ for a simpler version of the Root relation in which the upper bound on e is removed; let us call $R_{\mathsf{Root}'}$ this relation.

Below is an interactive ZK protocol for $R_{\mathsf{Root}'}$:

1. Prover computes a W such that $W^e = Acc$ and $C_W = WH^{r_2}, C_r = G^{r_2} H^{r_3}$ and sends to the verifier:
   $\mathcal{P} \rightarrow \mathcal{V} : C_W, C_r$
2. Prover and Verifier perform a protocol for the relation:
   $R((C_e, C_r, C_W, Acc), (e, r, r_2, r_3, \beta, \delta)) = 1$ iff

$$C_e = G^e H^r \wedge C_r = G^{r_2} H^{r_3} \wedge Acc = C_W^e \left(\frac{1}{H}\right)^{\beta} \wedge 1 = C_r^e \left(\frac{1}{H}\right)^{\delta} \left(\frac{1}{G}\right)^{\beta}$$

   Let λ_s be the size of the challenge space, λ_z be the statistical security parameter and μ the size of e.

   – Prover samples:

$$r_e \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right)$$
$$r_r, r_{r_2}, r_{r_3} \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$$
$$r_\beta, r_\delta \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right)$$

Root protocol

On common reference string $\mathsf{crs} = (\mathbb{Z}_N^*, G, H)$

$\mathsf{Prove}(\mathsf{crs}, (C_e, Acc), (e, r, w))$:

1. samples $r_2, r_3 \leftarrow\$ \left(-\lfloor N/4 \rfloor, \lfloor N/4 \rfloor\right)$ and computes $C_W \leftarrow WH^{r_2}, C_r \leftarrow G^{r_2}H^{r_3}$.
2. Computes the non-interactive version of the above protocol
   $r_e \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right), r_r, r_{r_2}, r_{r_3} \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$
   $r_\beta, r_\delta \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right)$
   $\alpha_1 \leftarrow G^{r_e}H^{r_r}, \alpha_2 \leftarrow G^{r_{r_2}}H^{r_{r_3}}, \alpha_3 \leftarrow C_W^{r_e}\left(\frac{1}{H}\right)^{r_\beta}, \alpha_4 \leftarrow C_r^{r_e}\left(\frac{1}{H}\right)^{r_\delta}\left(\frac{1}{G}\right)^{r_\beta}$
   $c \leftarrow \mathsf{H}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, C_e, Acc)$
   $s_e \leftarrow r_e - ce, s_r \leftarrow r_r - cr, s_{r_2} \leftarrow r_{r_2} - cr_2, s_{r_3} \leftarrow r_{r_3} - cr_{r_3}, s_\beta \leftarrow r_\beta - cer_2, s_\delta \leftarrow r_\delta - cer_3$

   Returns $\pi \leftarrow (C_W, C_r, \alpha_1, \alpha_2, \alpha_3, \alpha_4, s_e, s_r, s_{r_2}, s_{r_3}, s_\beta, s_\delta)$

$\mathsf{VerProof}(\mathsf{crs}, (C_e, Acc), \pi)$: returns 1 iff $\alpha_1 = C_e^c G^{s_e} H^{s_r} \wedge \alpha_2 = C_r^c G^{s_{r_2}} H^{s_{r_3}} \wedge \alpha_3 = Acc^c C_W^{s_e}\left(\frac{1}{H}\right)^{s_\beta} \wedge \alpha_4 = C_r^{s_e}\left(\frac{1}{H}\right)^{s_\delta}\left(\frac{1}{G}\right)^{s_\beta} \wedge s_e \in \left[-2^{\lambda+\lambda_s+\mu+1}, 2^{\lambda+\lambda_s+\mu+1}\right]$

Fig. 8 Description of the Root protocol

and computes:

$$\alpha_1 = G^{r_e}H^{r_r}, \quad \alpha_2 = G^{r_{r_2}}H^{r_{r_3}}, \quad \alpha_3 = C_W^{r_e}\left(\frac{1}{H}\right)^{r_\beta}, \quad \alpha_4 = C_r^{r_e}\left(\frac{1}{H}\right)^{r_\delta}\left(\frac{1}{G}\right)^{r_\beta}$$

$\mathcal{P} \rightarrow \mathcal{V} : (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$

– Verifier samples the challenge $c \leftarrow \{0,1\}^{\lambda_s}$ $\mathcal{V} \rightarrow \mathcal{P} : c$
– Prover computes the response:

$$s_e = r_e - ce$$
$$s_r = r_r - cr, \quad s_{r_2} = r_{r_2} - cr_2, \quad s_{r_3} = r_{r_3} - cr_{r_3}$$
$$s_\beta = r_\beta - cer_2, \quad s_\delta = r_\delta - cer_3$$

$\mathcal{P} \rightarrow \mathcal{V} : (s_e, s_r, s_{r_2}, s_{r_3}, s_\beta, s_\delta)$

– Verifier checks if:

$$\alpha_1 \stackrel{?}{=} C_e^c G^{s_e} H^{s_r}, \quad \alpha_2 \stackrel{?}{=} C_r^c G^{s_{r_2}} H^{s_{r_3}}, \quad \alpha_3 \stackrel{?}{=} Acc^c C_W^{s_e}\left(\frac{1}{H}\right)^{s_\beta},$$
$$\alpha_4 \stackrel{?}{=} C_r^{s_e}\left(\frac{1}{H}\right)^{s_\delta}\left(\frac{1}{G}\right)^{s_\beta}$$

Theorem 4.6 *Let $\mathbb{Z}_N^*$ be an RSA group where strong-RSA assumption holds, then the above protocol is a correct, knowledge sound and honest-verifier zero knowledge protocol for $R_{\mathsf{Root}'}$ (Fig. 8).*

The proof of the above is similar to the one of [16] where the more specific protocol was introduced, but implicitly was including a protocol for $R_{\mathsf{Root}'}$. Before proceeding to the proof we recall some properties related to RSA groups. First we expose two standard arguments. The first is that obtaining a multiple of $\phi(N)$ is equivalent to factoring N. This directly allows us to argue that for any $G \in \mathbb{Z}_N^*$, if one is able to find an $x \in \mathbb{Z}$ such that $G^x = 1 \pmod N$ then under the factoring assumption $x = 0$, otherwise x is a multiple of $\phi(N)$. Secondly, finding any non-trivial solution of the equation $\mu^2 = 1 \pmod N$ in $\mathbb{Z}_N^*$ (non-trivial means $\mu \neq \pm 1$) is equivalent to factoring N.

Remark 4.5 In 2017 Couteau et al. proved that in fact knowledge soundness for the protocol of opening an integer commitment can be reduced to (plain) RSA problem [25]. This could

be inherited to our protocol too. However, the relation itself assumes strong RSA's hardness, otherwise finding a root would be computable in polynomial time. Additionally, in the reduction to (plain) RSA, the extractor's probability of success is cubic, while in the reduction to strong RSA linear, in the adversary's probability of success.

Proposition 4.1 *Let $\mathbb{Z}_N^*$ be an RSA group with a modulus N and QR_N the corresponding group of quadratic residues modulo N.*

1. *Let $G, H \leftarrow\$ \mathsf{QR}_N$ two random generators of QR_N and a PPT adversary $\mathcal{A}$ outputting $\alpha, \beta \in \mathbb{Z}_N^*$ such that $G^\alpha H^\beta = 1$ then under the assumption that DLOG problem is hard in QR_N it holds that $\alpha = \beta = 0$.*
2. *Let $A, B \in \mathbb{Z}_N^*$ and a PPT adversary $\mathcal{A}$ outputting $x, y \in \mathbb{Z}_N^*$ such that $A^y = B^x$ and $y \mid x$ then under the assumption that factoring of N is hard it holds that $A = \pm B^{\frac{x}{y}}$.*

Proof 1. Since $G, H \in \mathsf{QR}_N$ there is an $x \in \mathbb{Z}_N^*$ such that $G = H^x \pmod N$ which leads to $H^{x\alpha+\beta} = 1$. As we discussed above under the assumption that factoring of N is hard, $x\alpha + \beta = 0$. If $\alpha \neq 0$ then $x \leftarrow -\frac{\beta}{\alpha}$ is a discrete logarithm of H, so assuming that DLOG is hard $\alpha = 0$. Similarly, there is an $y \in \mathbb{Z}_N^*$ such that $G^y = H \pmod N$ and with a similar argument we can conclude that $\beta = 0$.

2. We discern two cases, $y = \rho$ is odd or $y = 2^v\rho$ is even (for an odd ρ). In case y is odd then it is co-prime with $\phi(N) = p'q'$ (otherwise if $y = p'$ or $y = q'$ we would be able to factor N), so $y^{-1} \pmod{\phi(N)}$ exists and $A = B^{\frac{x}{y}}$. If $y = 2^v\rho$ then $\left(A^{-1}B^{\frac{x}{y}}\right)^y = 1 \Rightarrow \left(A^{-1}B^{\frac{x}{y}}\right)^{2^v\rho} = 1 \Rightarrow \left(A^{-1}B^{\frac{x}{y}}\right)^{2^v} = 1$. From the second fact that we discussed above under the factoring assumption $\left(A^{-1}B^{\frac{x}{y}}\right)^{2^{v-1}} = \pm 1$. However for $v > 1$ the left part of the equation is a quadratic residue so it cannot be -1, therefore $\left(A^{-1}B^{\frac{x}{y}}\right)^{2^{v-1}} = 1$. Using the same facts repeatedly we will eventually conclude that $\left(A^{-1}B^{\frac{x}{y}}\right)^2 = 1$, hence $A^{-1}B^{\frac{x}{y}} = \pm 1 \Rightarrow A = \pm B^{\frac{x}{y}}$.

□

Proof of Theorem 4.6 Correctness is straightforward. Honest-verifier zero knowledge can be shown with standard arguments used in Σ-protocols and the fact that the commitments to C_e, C_W, C_r are statistically hiding. That is the simulator $\mathcal{S}$ on input (C_e, Acc) samples $C_W^* \leftarrow\$ \mathbb{Z}_N^*$ and $C_r^* \leftarrow\$ \mathbb{Z}_N^*$. Then samples

$$s_e^* \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu} - 2^{\lambda_z+\mu}, 2^{\lambda_z+\lambda_s+\mu} + 2^{\lambda_z+\mu}\right),$$
$$s_r^*, s_{r_2}^*, s_{r_3}^* \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s} - \lfloor N/4 \rfloor 2^{\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s} + \lfloor N/4 \rfloor 2^{\lambda_s}\right),$$
$$s_\beta^*, s_\delta^* \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu} - \lfloor N/4 \rfloor 2^{\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu} + \lfloor N/4 \rfloor 2^{\lambda_s+\mu}\right).$$

Finally it samples $c^* \leftarrow\$ \{0,1\}^{\lambda_s}$. Then it sets $\alpha_1^* \leftarrow C_e^c G^{s_e} H^{s_r}$, $\alpha_2^* \leftarrow C_r^c G^{s_{r_2}} H^{s_{r_3}}$, $\alpha_3^* \leftarrow \mathsf{Acc}^c C_W^{s_e} \left(\frac{1}{H}\right)^{s_\beta}$ and $\alpha_4^* \stackrel{?}{=} C_r^{s_e} \left(\frac{1}{H}\right)^{s_\delta} \left(\frac{1}{G}\right)^{s_\beta}$. $\mathcal{S}$ outputs $\pi^* \leftarrow (C_W^*, C_r^*, \alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*, c^*, s_e^*, s_r^*, s_{r_2}^*, s_{r_3}^*, s_\beta^*, s_\delta^*)$. The distribution of π^* is identical to the one of a real proof π.

For the knowledge soundness, let an adversary of the knowledge soundness $\mathcal{A}$ that is able to convince the verifier $\mathcal{V}$ with a probability at least ϵ. We will construct an extractor $\mathcal{E}$ that extracts the witness $(e, r, r_2, r_3, \beta, \delta)$. Using rewinding $\mathcal{E}$ gets two accepted transcripts

$$(C_W, C_r, \alpha_1, \alpha_2, \alpha_3, \alpha_4, c, s_e, s_r, s_{r_2}, s_{r_3}, s_\beta, s_\delta)$$
$$\text{and } (C_W, C_r, \alpha_1, \alpha_2, \alpha_3, \alpha_4, c', s_e', s_r', s_{r_2}', s_{r_3}', s_\beta', s_\delta'),$$

on two different challenges c and c'. $\mathcal{E}$ aborts if it cannot get two such transcripts (abort1).

We denote $\Delta c := c' - c$, $\Delta s_e := s_e - s'_e$, $\Delta s_r := s_r - s'_r$, $\Delta s_{r_2} := s_{r_2} - s'_{r_2}$, $\Delta s_{r_3} := s_{r_3} - s'_{r_3}$, $\Delta s_\beta := s_\beta - s'_\beta$, $\Delta s_\delta := s_\delta - s'_\delta$ then

$$C_e^{\Delta c} = G^{\Delta s_e} H^{\Delta s_r}, C_r^{\Delta c} = G^{\Delta s_{r_2}} H^{\Delta s_{r_3}}, \mathsf{Acc}^{\Delta c} = C_W^{\Delta s_e} \left(\frac{1}{H}\right)^{\Delta s_\beta},$$

$$1 = C_r^{\Delta s_e} \left(\frac{1}{H}\right)^{\Delta s_\delta} \left(\frac{1}{G}\right)^{\Delta s_\beta}.$$

Define the (possibly rational) numbers $\hat{e} := \frac{\Delta s_e}{\Delta c}, \hat{r} := \frac{\Delta s_r}{\Delta c}, \hat{r_2} := \frac{\Delta s_{r_2}}{\Delta c}, \hat{r_3} := \frac{\Delta s_{r_3}}{\Delta c}$. In case Δc doesn't divide Δs_e and Δs_r, $\mathcal{E}$ aborts (abort $2a$). Similarly, in case Δc doesn't divide Δs_{r_2} and Δs_{r_3}, $\mathcal{E}$ aborts (abort $2b$). Therefore, since the above aborts didn't happen and according to second point of Proposition 4.1, $C_e = \pm G^{\hat{e}} H^{\hat{r}}$ and $C_r = \pm G^{\hat{r_2}} H^{\hat{r_3}}$.

Now if we replace C_r in the fourth equation we get $1 = (\pm 1)^{\Delta s_e} G^{\hat{r_2} \Delta s_e} H^{\hat{r_3} \Delta s_e} \left(\frac{1}{H}\right)^{\Delta s_\delta} \left(\frac{1}{G}\right)^{\Delta s_\beta}$ or $(\pm 1)^{\Delta s_e} G^{\hat{r_2} \Delta s_e - \Delta s_\beta} H^{\hat{r_3} \Delta s_e - \Delta s_\delta} = 1$. However, $(\pm 1)^{\Delta s_e} = 1$ otherwise if $(\pm 1)^{\Delta s_e} = -1$ then $-G^{\hat{r_2} \Delta s_e - \Delta s_\beta} H^{\hat{r_3} \Delta s_e - \Delta s_\delta}$ would be a non-quadratic residue (since G, H are both in QR_N and QR_N is closed under multiplication) equal to 1 which is a quadratic residue and this would be a contradiction, hence $G^{\hat{r_2} \Delta s_e - \Delta s_\beta} H^{\hat{r_3} \Delta s_e - \Delta s_\delta} = 1$. According to the first point of Proposition 4.1, under the factoring assumption $\hat{r_2} \Delta s_e - \Delta s_\beta = \hat{r_3} \Delta s_e - \Delta s_\delta = 0$, so $\hat{r_2} \Delta s_e = \Delta s_\beta$.

Finally we replace Δs_β in the third equation and we get $Acc^{\Delta c} = C_W^{\Delta s_e} \left(\frac{1}{H}\right)^{\hat{r_2} \Delta s_e} \Rightarrow Acc^{\Delta c} = \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\Delta s_e}$. As stated above Δc divides Δs_e so according to the second point of Proposition 4.1 $\mathsf{Acc} = \pm \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\frac{\Delta s_e}{\Delta c}} = \pm \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\hat{e}}$. We discern three cases:

- $\mathsf{Acc} = + \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\frac{\Delta s_e}{\Delta c}}$: Then $\mathcal{E}$ sets $\tilde{W} \leftarrow \frac{C_W}{H^{\hat{r_2}}}$ and $\tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$ $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ as above. It is clear that $\mathsf{Acc} = \tilde{W}^{\tilde{e}}$ and as stated above $C_e = G^{\tilde{e}} H^{\tilde{r}}$.
- $\mathsf{Acc} = - \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\frac{\Delta s_e}{\Delta c}}$ *and* $\frac{\Delta s_e}{\Delta c}$*odd*: Then $\mathcal{E}$ sets $\tilde{W} \leftarrow - \frac{C_W}{H^{\hat{r_2}}}$ and $\tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$ $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ as above. It is clear that $\mathsf{Acc} = \tilde{W}^{\tilde{e}}$ and as stated above $C_e = G^{\tilde{e}} H^{\tilde{r}}$.
- $\mathsf{Acc} = - \left(\frac{C_W}{H^{\hat{r_2}}}\right)^{\frac{\Delta s_e}{\Delta c}}$ *and* $\frac{\Delta s_e}{\Delta c}$*even*: this means that Acc is a non-quadratic residue, which is a contradiction since in the $R_{\mathsf{Root}'}$ relation we assume that $\mathsf{Acc} \in \mathsf{QR}_N$.

Finally the $\mathcal{E}$ outputs $(\tilde{e}, \tilde{r}, \tilde{W})$.

Now we show that the probability the extractor terminates with outputting a valid witness is $O(\epsilon)$. If the extractor does not abort then it clearly outputs a valid witness (under factoring assumption). For the first abort, with a standard argument it can be shown that the extractor is able to extract two accepting transcripts with probability $O(\epsilon)$ (for the probabilistic analysis we refer to [31]). Thus $Pr[\text{abort1}] = 1 - O(\epsilon)$. For the second type of aborts (abort $2a$ and abort $2b$), they happen with negligible probability under the strong RSA assumption. For the details see Lemma 4.2 below, which was proven in [31]. Putting them together the probability of success of $\mathcal{E}$ is at least $O(\epsilon) - \mathsf{negl}(\lambda_s)$. □

Lemma 4.2 ([31]) *Given that* abort $2a$ *occurs a PPT adversary $\mathcal{B}$ can solve the strong RSA problem with probability at least* $\frac{1}{2} - 2^{-\lambda_s}$.

From the above we get $Pr[\mathcal{B} \text{ solves } sRSA] \geq \left(\frac{1}{2} - 2^{-\lambda_s}\right) Pr[\mathsf{abort}\, 2a]$, so we conclude to $Pr[\mathsf{abort}\, 2a] \leq \frac{1}{\frac{1}{2}-2^{-\lambda_s}} Pr[\mathcal{B} \text{ solves } sRSA] = \mathsf{negl}(\lambda_s)$. The same lemma holds for $\mathsf{abort}\, 2b$.

Notice in the above protocol that

$$\begin{aligned}
&-2^{\lambda_z+\lambda_s+\mu} - 2^{\lambda_s+\mu} \leq s_e \leq 2^{\lambda_z+\lambda_s+\mu} + 2^{\lambda_s+\mu} \Rightarrow \\
&-2^{\lambda_z+\lambda_s+\mu+1} \leq s_e \leq 2^{\lambda_z+\lambda_s+\mu+1} \Rightarrow \\
&-2^{\lambda_z+\lambda_s+\mu+2} \leq \Delta s_e \leq 2^{\lambda_z+\lambda_s+\mu+2} \Rightarrow \\
&-2^{\lambda_z+\lambda_s+\mu+2} \leq \hat{e} \leq 2^{\lambda_z+\lambda_s+\mu+2},
\end{aligned}$$

so if we impose an additional verification check of honest s_e size, i.e., $s_e \in \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right]$, we get that $|\hat{e}| \leq 2^{\lambda_z+\lambda_s+\mu+2}$. The verifier performs an extra range check $s_e \stackrel{?}{\in} \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right]$ and the resulting protocol is the $\mathsf{CP}_{\mathsf{Root}}$ that except for proving of knowledge of an e-th root also provides a bound for the size of $|e|$:

$$\begin{aligned}
R_{\mathsf{Root}}((C_e, Acc, \mu), (e, r, W)) = 1 \text{ iff } C_e = \pm G^e H^r \pmod N \wedge W^e = Acc \\
\pmod N \wedge |e| < 2^{\lambda_z+\lambda_s+\mu+2}.
\end{aligned}$$

4.5.2 Protocol $\mathsf{CP}_{\mathsf{modEq}}$

Below we describe the public-coin ZK protocol for R_{modEq}. In Fig. 9 we summarize the corresponding NIZK obtained after applying the Fiat–Shamir transform to it.

1. Prover samples:

$$\begin{aligned}
r_e &\leftarrow \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right) \\
r_r &\leftarrow \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right) \\
r_{r_q} &\leftarrow \mathbb{Z}_q,
\end{aligned}$$

   and computes:

$$\alpha_1 = G^{r_e} H^{r_r}, \quad \alpha_2 = g^{r_e} \pmod p h^{r_{r_q}}.$$

   $\underline{\mathcal{P} \rightarrow \mathcal{V}}$: (α_1, α_2).
2. Verifier samples the challenge $c \leftarrow \{0,1\}^{\lambda_s}$.para $\underline{\mathcal{V} \rightarrow \mathcal{P} c}$.
3. Prover computes the response:

$$\begin{aligned}
s_e &= r_e - ce \\
s_r &= r_r - cr \\
s_{r_q} &= r_{r_q} - cr_q \pmod q.
\end{aligned}$$

   $\underline{\mathcal{P} \rightarrow \mathcal{V}}$: (s_e, s_r, s_{r_q}).
4. Verifier checks if:

$$\alpha_1 \stackrel{?}{=} \pm C_e^c G^{s_e} H^{s_r} \pmod N, \alpha_2 \stackrel{?}{=} c_{e_q}^c g^{s_e} \pmod q h^{s_{r_q}}.$$

Theorem 4.7 *Let $\mathbb{Z}_N^*$ be an RSA group where strong-RSA assumption holds and $\mathbb{G}$ be a prime order group where DLOG assumption holds then the above protocol is a correct, knowledge sound and honest-verifier zero knowledge protocol for R_{modEq}.*

The proof is quite simple and is omitted.

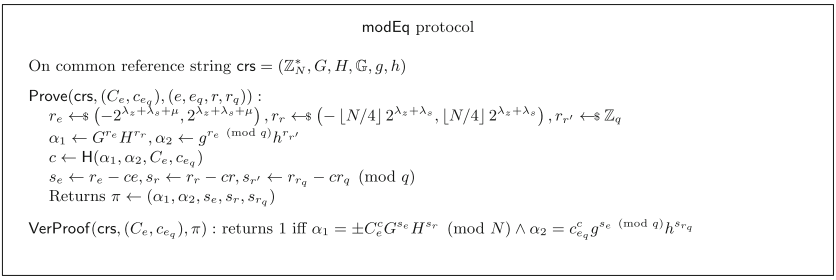


Fig. 9 Description of the modEq protocol

4.6 Instantiations

We discuss the possible instantiations of our schemes $\mathsf{MemCP}_{\mathsf{RSA}}$ and $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ that can be obtained by looking at applications' constraints and security parameters constraints. *Parameters for* $d\mu + 2 \leq \nu$ *and* $\mu \leq \nu - 2$. First we analyze possible parameters that satisfy the conditions $d\mu + 2 \leq \nu \wedge \mu \leq \nu - 2$ that is used in Theorems 4.1 and 4.2; we recall $d = 1 + \lfloor \frac{\lambda_z+\lambda_s+2}{\mu} \rfloor$, where λ_z and λ_s are statistical security parameters for zero-knowledge and soundness respectively of $\mathsf{CP}_{\mathsf{Root}}$.

If the prime order group $\mathbb{G}_q$ is instantiated with (pairing-friendly) elliptic curves, then the bitsize ν of its order must be at least 2λ. And recall that for correctness we need $\mu < \nu$.

Considering these constraints, one way to satisfy $d\mu + 2 \leq \nu$ is to choose μ such that $\nu - 1 > \mu > \lambda_z + \lambda_s + 2$. More specifically, a choice that maximizes security is $\nu = 2\lambda$, $\mu = 2\lambda - 2$ and $\lambda_z = \lambda - 3, \lambda_s = \lambda - 2$. For the case of the $\mathsf{MemCP}_{\mathsf{RSA}}$ scheme, this choice yields an instantiation with nearly λ bits of security and where the function H does not necessarily need to be a random oracle (yet it must be collision resistant).

Because of the constraint $\mu > \lambda_z + \lambda_s + 2$, we the choice above implies the use of large primes. This would be anyway the case if one instantiates the scheme with a collision-resistant hash function H (e.g., SHA256 or SHA3), e.g., because set elements are quite arbitrary. If on the other hand, one could support more specific set elements, one could use instead a deterministic map-to-primes or even use our scheme $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ in which set elements themselves are primes. In this case one may wonder if it is possible to choose values of μ smaller than 2λ; for example $\mu \approx 30, 60, 80$. The answer is positive although the characterization of such μ's require an involved analysis.

Let us fix $\nu = 2\lambda$, and say that the statistical security parameters λ_z, λ_s are such that $\lambda_z + \lambda_s + 2 = 2\lambda - 2 - c$ for some constant c (for example $c = 4$ if $\lambda_z = \lambda_s = \lambda - 4$). We are essentially looking for μ such that

$$\begin{aligned} &\mu \leq 2\lambda - 2 - c \text{ and } \mu + \mu \left\lfloor \frac{2\lambda - 2}{\mu} - \frac{c}{\mu} \right\rfloor \leq 2\lambda - 2 \\ \iff\ &\mu \leq 2\lambda - 2 - c \text{ and } \left\lfloor \frac{2\lambda - 2}{\mu} - \frac{c}{\mu} \right\rfloor \leq \frac{2\lambda - 2}{\mu} - 1. \end{aligned}$$

From the fact $x \mod y = x - y\lfloor \frac{x}{y} \rfloor$, we can reduce the above inequality into

$$\mu \leq 2\lambda - 2 - c \text{ and } 2\lambda - 2 - c \bmod \mu \geq \mu - c,$$

that can admit solutions for $c \geq 2$.

For instance, if $\lambda = 128$ and $c = 4$, then we get several options for μ, e.g., $\mu = 32, 42, 63, 84, 126, 127$.

Parameters for $d\mu + 2 > \nu$. This case concerns only $\mathsf{MemCP}_{\mathsf{RSA}}$ and Theorem 4.2 in particular. In this case, if one aims at maximizing security, say to get a scheme with λ-bits of security, then would have to set $\mu \approx 2\lambda$ for collision resistance, and consequently select the prime order group so that $\nu \geq 3\lambda$. This choice however is costly in terms of performance since the efficiency of all protocols that work in the prime order group degrades.

5 A CP-SNARK for set non-membership with short parameters

Here we describe two CP-SNARKs for set non-membership that work in a setting identical to the one of Sect. 4. Namely, the set is committed using an RSA accumulator, and the element (that one wants to prove not to belong to the set) is committed using a Pedersen commitment scheme. As in the previous section, we propose two protocols for non-membership, called $\mathsf{NonMemCP}_{\mathsf{RSA}}$ and $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$, in complete analogy to $\mathsf{MemCP}_{\mathsf{RSA}}$ and $\mathsf{MemCP}_{\mathsf{RSAPrm}}$. In the former, the elements of the set are arbitrary bit-strings of length η, $\mathcal{D}_{\mathsf{elm}} = \{0,1\}^{\eta}$, while in the latter the elements are primes of length μ. The schemes are fully described in Figs. 10 and 11.

5.1 An high-level overview of the constructions

The main idea of $\mathsf{NonMemCP}_{\mathsf{RSA}}$ is similar to the one of the corresponding membership protocol, $\mathsf{MemCP}_{\mathsf{RSA}}$. It uses in the same modular way the modEq and HashEq protocols. The only difference lies in the third protocol: instead of using Root it uses a new protocol $\mathsf{Coprime}$. In a similar manner, $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$ uses modEq, Range and $\mathsf{Coprime}$.

Let us explain the need of the $\mathsf{Coprime}$ protocol and what it does. First, recall how a non-membership proof is computed in RSA Accumulators [50]. Let P be a set of primes to be accumulated and prod the corresponding product. For any prime element $e \notin P$ it holds that $\gcd(e, \mathsf{prod}) = 1$, while for any member $e \in P$ it is $\gcd(e, \mathsf{prod}) = e \neq 1$. Thus, proving that $\gcd(e, \mathsf{prod}) = 1$ would exhibit non-membership of e in P. Recall, also, that using the extended Euclidean algorithm one can efficiently compute coefficients (a, b) such that $a \cdot e + b \cdot \mathsf{prod} = \gcd(e, \mathsf{prod})$. A non-membership proof for an element e w.r.t. an accumulator $\mathsf{Acc} = G^{\mathsf{prod}}$ consists of a pair $(D = G^a, b)$, where a, b are such that $a \cdot e + b \cdot \mathsf{prod} = 1$. The verification is $D^e \mathsf{Acc}^b = G$, which ensures that e and prod are coprime, i.e. $\gcd(e, \mathsf{prod}) = 1$. Therefore, the goal of the $\mathsf{Coprime}$ protocol is to prove knowledge of an element e committed in an integer commitment C_e that satisfies this relation. A more formal definition of $\mathsf{Coprime}$ is given below and an instantiation of this protocol is in Sect. 5.4.

5.2 Argument of knowledge for a coprime element

We make use of a non-interactive argument of knowledge of a non-membership witness of an element such that the verification equation explained above holds. More formally $\mathsf{CP}_{\mathsf{Coprime}}$, is a NIZK for the relation: $R_{\mathsf{Coprime}} : (\mathbb{Z}_N^* \times \mathsf{QR}_N) \times (\mathbb{Z} \times \mathbb{Z} \times \mathsf{QR}_N \times \mathbb{Z})$ defined as

$R_{\mathsf{Coprime}}((C_e, \mathsf{Acc}), (e, r, D, b)) = 1$ iff

$$C_e = \pm G^e H^r \bmod N \ \wedge \ D^e \mathsf{Acc}^b = G \wedge |e| < 2^{\lambda_z + \lambda_s + \mu + 2}.$$

$\mathsf{KeyGen}(\mathsf{ck}, R^{\in})$: parse $\mathsf{ck} := ((N, G, \mathsf{H}_{\mathsf{prime}}), (\mathbb{G}_q, g, h))$ as the commitment keys of $\mathsf{SetCom}_{\mathsf{RSA}}$ and PedCom respectively. Sample a random generator H.
Generate $\mathsf{crs}_{\mathsf{HashEq}} \leftarrow\!\$\ \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{KeyGen}((\mathbb{G}_q, g, h), R_{\mathsf{HashEq}})$, a crs for $\mathsf{CP}_{\mathsf{HashEq}}$.
Return $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{HashEq}})$.
Given crs, one can define $\mathsf{crs}_{\mathsf{Coprime}} := (N, G, H)$, $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$.

$\mathsf{Prove}(\mathsf{crs}, (C_U, c_u), (U, u), (\varnothing, r_u))$: $e \leftarrow \mathsf{H}_{\mathsf{prime}}(u) = (1|\mathsf{H}(u, j))$, $(c_e, r_q) \leftarrow \mathsf{Com}_1.\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_q, e)$.
$(C_e, r) \leftarrow \mathsf{IntCom}.\mathsf{Commit}((G, H), e)$; $P \leftarrow \{\mathsf{H}_{\mathsf{prime}}(u) : u \in U\}$, compute a, b s.t. $a \cdot e + b \cdot \prod_{e_i \in P} e_i = 1$ and set $D = G^a$.
$\pi_{\mathsf{Coprime}} \leftarrow \mathsf{CP}_{\mathsf{Coprime}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Coprime}}, (C_e, C_U, \mu), (e, r, D, b))$
$\pi_{\mathsf{modEq}} \leftarrow \mathsf{CP}_{\mathsf{modEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), (e, e, r, r_q))$
$\pi_{\mathsf{HashEq}} \leftarrow \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{HashEq}}, (c_e, c_u), (e, u), (r_q, r_u), j)$
Return $\pi := (C_e, c_e, \pi_{\mathsf{Root}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{HashEq}})$.

$\mathsf{VerProof}(\mathsf{crs}, (C_U, c_u), \pi)$: Return 1 iff $\mathsf{CP}_{\mathsf{Root}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Coprime}}, (C_e, C_U, \mu), \pi_{\mathsf{Coprime}}) = 1 \wedge$
$\mathsf{CP}_{\mathsf{modEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), \pi_{\mathsf{modEq}}) = 1 \wedge \mathsf{CP}_{\mathsf{HashEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{HashEq}}, (c_e, c_u), \pi_{\mathsf{HashEq}}) = 1$.

Fig. 10 $\mathsf{NonMemCP}_{\mathsf{RSA}}$ CP-SNARK for set non-membership

$\mathsf{KeyGen}(\mathsf{ck}, R^{\in})$: parse $\mathsf{ck} := ((N, G, \mathsf{H}_{\mathsf{prime}}), (\mathbb{G}_q, g, h))$ as the commitment keys of $\mathsf{SetCom}_{\mathsf{RSA}'}$ and PedCom respectively. Sample a random generator H.
Generate $\mathsf{crs}_{\mathsf{Range}} \leftarrow\!\$\ \mathsf{CP}_{\mathsf{Range}}.\mathsf{KeyGen}((\mathbb{G}_q, g, h), R_{\mathsf{Range}})$, a crs for $\mathsf{CP}_{\mathsf{Range}}$.
Return $\mathsf{crs} := (N, G, H, \mathsf{H}_{\mathsf{prime}}, \mathbb{G}_q, g, h, \mathsf{crs}_{\mathsf{Range}})$.
Given crs, one can define $\mathsf{crs}_{\mathsf{Coprime}} := (N, G, H)$, $\mathsf{crs}_{\mathsf{modEq}} := (N, G, H, \mathbb{G}_q, g, h)$.

$\mathsf{Prove}(\mathsf{crs}, (C_P, c_e), (P, e), (\varnothing, r_e))$: $(C_e, r) \leftarrow \mathsf{IntCom}.\mathsf{Commit}((G, H), e)$; , compute a, b s.t. $a \cdot e + b \cdot \prod_{e_i \in P} e_i = 1$ and set $D = G^a$.
$\pi_{\mathsf{Coprime}} \leftarrow \mathsf{CP}_{\mathsf{Coprime}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Coprime}}, (C_e, C_P, \mu), (e, r, D, b))$
$\pi_{\mathsf{modEq}} \leftarrow \mathsf{CP}_{\mathsf{modEq}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), (e, e, r, r_q))$
$\pi_{\mathsf{Range}} \leftarrow \mathsf{CP}_{\mathsf{Range}}.\mathsf{Prove}(\mathsf{crs}_{\mathsf{Range}}, (2^{\mu-1}, 2^{\mu}), c_e, e, r_q)$
Return $\pi := (C_e, c_e, \pi_{\mathsf{Coprime}}, \pi_{\mathsf{modEq}}, \pi_{\mathsf{Range}})$.

$\mathsf{VerProof}(\mathsf{crs}, (C_P, c_e), \pi)$: Return 1 iff $\mathsf{CP}_{\mathsf{Coprime}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Coprime}}, (C_e, C_P, \mu), \pi_{\mathsf{Coprime}}) = 1 \wedge$
$\mathsf{CP}_{\mathsf{modEq}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{modEq}}, (C_e, c_e), \pi_{\mathsf{modEq}}) = 1 \wedge \mathsf{CP}_{\mathsf{Range}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{Range}}, c_e, \pi_{\mathsf{Range}}) = 1$.

Fig. 11 $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$ CP-SNARK for set non-membership

We propose an instantiation of a protocol for the above relation in the Sect. 5.4.

5.3 Our constructions of $\mathsf{NonMemCP}_{\mathsf{RSA}}$ and $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$

In Figs. 10 and 11 we give a full description of the schemes.

The security of these schemes follow very closely the one of the corresponding membership schemes given in Sect. 4. Below we give the Theorems that state their security. The proofs are omitted since they are almost identical to the corresponding proofs for the membership schemes.

Theorem 5.1 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}}$ *and* IntCom *be computationally binding commitments,* $\mathsf{CP}_{\mathsf{Coprime}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{HashEq}}$ *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption hold, and that* H *is collision resistant.*

If $d\mu + 2 \leq \nu$, $\lambda_s + 1 < \mu$ *and* $\lambda_s < \log(N)/2$ *then* $\mathsf{NonMemCP}_{\mathsf{RSA}}$ *is knowledge-sound with partial opening of the set commitments* C_U.

Theorem 5.2 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}}$ *and* IntCom *be computationally binding commitments,* $\mathsf{CP}_{\mathsf{Coprime}}$, $\mathsf{CP}_{\mathsf{modEq}}$ *and* $\mathsf{CP}_{\mathsf{HashEq}}$ *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption hold, and that* H *is collision resistant.*

If $d\mu+2 > \nu, \lambda_s+1 < \mu, \lambda_s < \log(N)/2, d = O(1)$ *is a small constant,* $2^{\mu-\nu} \in \mathsf{negl}(\lambda)$ *and* H *is modeled as a random oracle, then* $\mathsf{NonMem}\mathcal{CP}_{\mathsf{RSA}}$ *is knowledge-sound with partial opening of the set commitments* C_U.

Theorem 5.3 *Let* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}'}$ *and* IntCom *be computationally binding commitments,* $\mathcal{CP}_{\mathsf{Coprime}}$, $\mathcal{CP}_{\mathsf{modEq}}$ *and* $\mathcal{CP}_{\mathsf{Range}}$ *be knowledge-sound NIZK arguments, and assume that the Strong RSA assumption hold. If* $d\mu + 2 \leq \nu$, $\lambda_s + 1 < \mu$ *and* $\lambda_s < \log(N)/2$ *then* $\mathsf{NonMem}\mathcal{CP}_{\mathsf{RSAPrm}}$ *is knowledge-sound with partial opening of the set commitments* c_P. *Furthermore, if* PedCom, $\mathsf{SetCom}_{\mathsf{RSA}'}$ *and* IntCom *are statistically hiding commitments, and* $\mathcal{CP}_{\mathsf{Coprime}}$, $\mathcal{CP}_{\mathsf{modEq}}$ *and* $\mathcal{CP}_{\mathsf{Range}}$ *be zero-knowledge, then* $\mathsf{NonMem}\mathcal{CP}_{\mathsf{RSAPrm}}$ *is zero-knowledge.*

5.4 Proposed instantiations of protocol for R_{Coprime}

Below we propose an interactive ZK protocol for R_{Coprime}. As the relation indicates, we need to prove knowledge of (D, b) such that $D^e\mathsf{Acc}^b = G$, for a committed e. Proving opening of C_e to e is straightforward, so the main challenge is to prove the non-membership equation. For this the prover should send D and Acc^b to the verifier so that she can check that $D^e\mathsf{Acc}^b = G$ herself. Of course, there are two caveats. The first one is that D and Acc^b cannot be sent in the plain as we require zero-knowledge; we solve this by sending them in a hiding manner, i.e., $C_a = DH^{r_a}$ and $C_B = \mathsf{Acc}^b H^{\rho_B}$ for random values r_a, ρ_B. Consequently, the verification now should work with the hiding elements. Secondly, the verifier should be ensured that Acc^b is indeed an exponentiation of Acc with a known (to the prover) value b, otherwise soundness can be broken. More specifically we require extraction of b, ρ_B such that $C_B = \mathsf{Acc}^b H^{\rho_B}$. This is done using the partial opening of Acc to the set represented by prod, i.e., the protocol assumes that $\mathsf{Acc} = G^{\mathsf{prod}}$ is a common knowledge.

Below we present our protocol in full details.

1. Prover computes $C_a = DH^{r_a}, C_{r_a} = G^{r_a}H^{r'_a}, C_B = Acc^b H^{\rho_B}, C_{\rho_B} = G^{\rho_B}H^{\rho'_B}$ and sends to the verifier:
   $\underline{\mathcal{P} \to \mathcal{V}}: C_a, C_{r_a}, C_B, C_{\rho_B}$.
2. Prover and Verifier perform a protocol for the relation: $R((\mathsf{Acc}, C_e, C_a, C_{r_a}, C_B, C_{\rho_B}), (e, b, r, r_a, r'_a, \rho_B, \rho'_B, \beta, \delta)) = 1$ iff

$$C_e = G^e H^r \wedge C_r = G^{r_2} H^{r_3} \wedge Acc = C_W^e \left(\frac{1}{H}\right)^{\beta} \wedge 1 = C_r^e \left(\frac{1}{H}\right)^{\delta} \left(\frac{1}{G}\right)^{\beta}.$$

   Let λ_s be the size of the challenge space, λ_z be the statistical security parameter and μ the size of e.

   – Prover samples:

$$r_b, r_e \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right)$$
$$r_{\rho_B}, r_r, r_{r_a}, r_{r'_a}, r_{\rho'_B} \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$$
$$r_\beta, r_\delta \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right),$$

   and computes:

$$\alpha_2 = \mathsf{Acc}^{r_b} H^{r_{\rho_B}}, \quad \alpha_3 = G^{r_e} H^{r_r}, \quad \alpha_4 = G^{r_{ra}} H^{r_{r'_a}},$$
$$\alpha_5 = C_a^{r_e} H^{r_\beta}, \quad \alpha_6 = C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta}, \quad \alpha_7 = G^{r_{\rho_B}} H^{r_{\rho'_B}}.$$

   $\underline{\mathcal{P} \to \mathcal{V}}: (\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7)$

Coprime protocol

On common reference string $\mathsf{crs} = (\mathbb{Z}_N^*, G, H)$

$\mathsf{Prove}(\mathsf{crs}, (C_e, \mathsf{Acc}), (e, r, (D, b)))$:

1. samples $r_a, r_{a'}, \rho_B, \rho_{B'} \xleftarrow{\$} (-\lfloor N/4 \rfloor, \lfloor N/4 \rfloor)$ and computes $C_a = DH^{r_a}, C_{r_a} = G^{r_a} H^{r'_a}, C_B = Acc^b H^{\rho_B}, C_{\rho_B} = G^{\rho_B} H^{\rho'_B}$.
2. Computes the non-interactive version of the above protocol
   $r_b, r_e \xleftarrow{\$} \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right)$
   $r_{\rho_B}, r_r, r_{r_a}, r_{r'_a}, r_{\rho'_B} \xleftarrow{\$} \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$
   $r_\beta, r_\delta \xleftarrow{\$} \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right)$
   $\alpha_2 = \mathsf{Acc}^{r_b} H^{r_{\rho_B}}, \alpha_3 = G^{r_e} H^{r_r}, \alpha_4 = G^{r_{r_a}} H^{r'_{r_a}}, \alpha_5 = C_a^{r_e} H^{r_\beta}, \alpha_6 = C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta}, \alpha_7 = G^{r_{\rho_B}} H^{r_{\rho'_B}}$
   $c \leftarrow \mathsf{H}(\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, C_e, Acc)$
   $s_b = r_b - cb, s_e = r_e - ce, s_{\rho_B} = r_{\rho_B} - c\rho_B, s_r = r_r - cr, s_{r_a} = r_{r_a} - cr_a, s_{r'_a} = r_{r'_a} - cr_{r_a}, s_{\rho'_B} = r_{\rho'_B} - c\rho'_B, s_\beta = r_\beta + c(er_a + \rho_B), s_\delta = r_\delta + c(er'_a + \rho'_B)$
   Returns $\pi \leftarrow (C_a, C_{r_a}, C_B, C_{\rho_B}, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, s_b, s_e, s_{r_b}, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta)$

$\mathsf{VerProof}(\mathsf{crs}, (C_e, \mathsf{Acc}), \pi)$: returns 1 iff $\alpha_2 = C_B^c \mathsf{Acc}^{s_b} H^{s_{\rho_B}} \wedge \alpha_3 = C_e^c G^{s_e} H^{s_r} \wedge \alpha_4 = C_{r_a}^c G^{s_{r_a}} H^{s_{r'_a}} \wedge \alpha_5 = C_a^{s_e} H^{s_\beta} G^c C_B^{-c} \wedge \alpha_6 = C_{r_a}^{s_e} H^{s_\delta} G^{s_\beta} C_{\rho_B}^{-c} \wedge \alpha_7 = C_{\rho_B}^c G^{s_{\rho_B}} H^{s_{\rho'_B}} \wedge \wedge s_e \in \left[-2^{\lambda+\lambda_s+\mu+1}, 2^{\lambda+\lambda_s+\mu+1}\right]$

Fig. 12 Description of the Coprime protocol

- Verifier samples the challenge $c \leftarrow \{0,1\}^{\lambda_s}$ $\underline{\mathcal{V} \rightarrow \mathcal{P}}$: c.
- Prover computes the response:

  $s_b = r_b - cb, s_e = r_e - ce$

  $s_{\rho_B} = r_{\rho_B} - c\rho_B, s_r = r_r - cr, s_{r_a} = r_{r_a} - cr_a, s_{r'_a} = r_{r'_a} - cr'_a, s_{\rho'_B} = r_{\rho'_B} - c\rho'_B$

  $s_\beta = r_\beta + c(er_a + \rho_B), \quad s_\delta = r_\delta + c(er'_a + \rho'_B).$

  $\underline{\mathcal{P} \rightarrow \mathcal{V}}$: $(s_b, s_e, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta)$
- Verifier checks if:

$$\alpha_2 \stackrel{?}{=} C_B^c \mathsf{Acc}^{s_b} H^{s_{\rho_B}}, \quad \alpha_3 \stackrel{?}{=} C_e^c G^{s_e} H^{s_r}, \quad \alpha_4 \stackrel{?}{=} C_{r_a}^c G^{s_{r_a}} H^{s_{r'_a}},$$
$$\alpha_5 \stackrel{?}{=} C_a^{s_e} H^{s_\beta} G^c C_B^{-c}, \quad \alpha_6 \stackrel{?}{=} C_{r_a}^{s_e} H^{s_\delta} G^{s_\beta} C_{\rho_B}^{-c}, \quad \alpha_7 \stackrel{?}{=} C_{\rho_B}^c G^{s_{\rho_B}} H^{s_{\rho'_B}},$$
$$s_e \stackrel{?}{\in} \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right].$$

5.4.1 Correctness

Here we show the correctness of the protocol (Fig. 12).

$$\begin{aligned}
\alpha_2 &= \mathsf{Acc}^{r_b} H^{r_{\rho_B}} = \mathsf{Acc}^{s_b+cb} H^{s_{\rho_B}+c\rho_B} = \mathsf{Acc}^{s_{r_b}} H^{s_{\rho_B}} (\mathsf{Acc}^b H^{\rho_B})^c \\
&= \mathsf{Acc}^{s_b} H^{s_{\rho_B}} C_B^c \\
\alpha_3 &= G^{r_e} H^{r_r} = G^{s_e+ce} H^{s_r+cr} = G^{s_e} H^{s_r} (G^e H^r)^c \\
&= G^{s_e} H^{s_r} C_e^c \\
\alpha_4 &= G^{r_{ra}} H^{r_{r'_a}} = G^{s_{ra}+cr_a} H^{s_{r'_a}+cr'_a} = G^{s_{ra}} H^{s_{r'_a}} (G^{r_a} H^{r'_a})^c \\
&= G^{s_{ra}} H^{s_{r'_a}} C_{r_a}^c \\
\alpha_5 &= C_a^{r_e} H^{r_\beta} = C_a^{s_e+ce} H^{s_\beta - c(er_a+\rho_B)} = C_a^{s_e} H^{s_\beta} (D^e H^{er_a})^c H^{-c(er_a+\rho_B)} \\
&= C_a^{s_e} H^{s_\beta} (D^e H^{-\rho_B})^c = C_a^{s_e} H^{s_\beta} (G Acc^{-b} H^{-\rho_B})^c = \\
&= C_a^{s_e} H^{s_\beta} G^c C_B^{-c} \\
\alpha_6 &= C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta} = C_{r_a}^{s_e+ce} G^{s_\beta - c(er_a+\rho_B)} H^{s_\delta - c(er'_a+\rho'_B)}
\end{aligned}$$

$$= C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} (G^{r_a} H^{r'_a})^{ce} G^{-c(er_a+\rho_B)} H^{-c(er'_a+\rho'_B)} = C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} G^{-c\rho_B} H^{-c\rho'_B}$$
$$= C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} C_{\rho_B}^{-c}$$

$$\alpha_7 = G^{r\rho_B} H^{r\rho'_B} = G^{s_{\rho_B}+c\rho_B} H^{s_{\rho'_B}+c\rho'_B} = G^{s_{\rho_B}} H^{s_{\rho'_B}} (G^{\rho_B} H^{\rho'_B})^c$$
$$= G^{s_{\rho_B}} H^{s_{\rho'_B}} C_{\rho_B}^c$$

.

5.4.2 Security

Security of our scheme holds with the partial opening of Acc, i.e., when it is ensured outside the protocol that Acc is a valid commitment of the set. The proof is similar to the one of Theorem 4.6. The main technical difference is in the extraction of the opening of C_B, because Acc is not a random generator sampled at the setup phase. However, from partial opening we know that it is $\mathsf{Acc} = G^{\mathsf{prod}}$ for a random generator G. This will allow us to state an alternative to Lemma 4.2 to justify the extraction of the opening of C_B.

Theorem 5.4 *Let $\mathbb{Z}_N^*$ be an RSA group where strong-RSA assumption holds, then the above protocol is honest-verifier zero knowledge protocol and, also, if $\lambda_s + 1 < \mu$ and $\lambda_s < \log(N)/2$, is knowledge sound with partial opening of* Acc *for* R_{Coprime}.

Proof Zero-Knowledge can be proven with standard techniques, similar to the ones in the proof of Theorem 4.6 and is therefore omitted.

For the knowledge soundness, let an adversary of the knowledge soundness $\mathcal{A}$ that is able to convince the verifier $\mathcal{V}$ with a probability at least ϵ. We will construct an extractor $\mathcal{E}$ that extracts the witness $(e, r, r_2, r_3, \beta, \delta)$. Using rewinding $\mathcal{E}$ gets two accepted transcripts

$$(C_a, C_{r_a}, C_B, C_{\rho_B}, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, c, s_b, s_e, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta)$$
$$(C_a, C_{r_a}, C_B, C_{\rho_B}, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, c', s'_b, s'_e, s'_{\rho_B}, s'_r, s'_{r_a}, s'_{r'_a}, s'_{\rho'_B}, s'_\beta, s'_\delta),$$

on two different challenges c and c'. $\mathcal{E}$ aborts if it cannot get two such transcripts (abort1).

We denote $\Delta c := c' - c$, $\Delta s_b := s_b - s'_b$, $\Delta s_e := s_e - s'_e$, $\Delta s_{\rho_B} := s_{\rho_B} - s'_{\rho_B}$, $\Delta s_r := s_r - s'_r$, $\Delta s_{r_a} := s_{r_a} - s'_{r_a}$, $\Delta s_{r'_a} := s_{r'_a} - s'_{r'_a}$, $\Delta s_{\rho'_B} := s_{\rho'_B} - s'_{\rho'_B}$, $\Delta s_\beta := s_\beta - s'_\beta$, $\Delta s_\delta := s_\delta - s'_\delta$ then

$$C_B^{\Delta c} = \mathsf{Acc}^{\Delta s_b} H^{\Delta s_{\rho_B}} \Rightarrow C_B = \pm \mathsf{Acc}^{\hat{b}} H^{\hat{\rho_B}}, \tag{1}$$
$$C_e^{\Delta c} = G^{\Delta s_e} H^{\Delta s_r} \Rightarrow C_e = \pm G^{\hat{e}} H^{\hat{r}}, \tag{2}$$
$$C_{r_a}^{\Delta c} = G^{\Delta s_{ra}} H^{\Delta s_{r'_a}} \Rightarrow C_{r_a} = \pm G^{\hat{r_a}} H^{\hat{r'_a}}, \tag{3}$$
$$1 = C_a^{\Delta s_e} H^{\Delta s_\beta} G^{-\Delta c} C_B^{\Delta c}, \tag{4}$$
$$1 = C_{r_a}^{\Delta s_e} H^{\Delta s_\delta} G^{\Delta s_\beta} C_{\rho_B}^{\Delta c}, \tag{5}$$
$$C_{\rho_B}^{\Delta c} = G^{\Delta s_{\rho_B}} H^{\Delta s_{\rho'_B}} \Rightarrow C_{\rho_B} = \pm G^{\hat{\rho_B}} H^{\hat{\rho'_B}}, \tag{6}$$

define the (possibly rational) numbers $\hat{b} := \frac{\Delta s_b}{\Delta c}$, $\hat{e} := \frac{\Delta s_e}{\Delta c}$, $\hat{r} := \frac{\Delta s_r}{\Delta c}$, $\hat{r_a} := \frac{\Delta s_{ra}}{\Delta c}$, $\hat{r'_a} := \frac{\Delta s_{r'_a}}{\Delta c}$, $\hat{\rho_B} := \frac{\Delta s_{\rho_B}}{\Delta c}$, $\hat{\rho'_B} := \frac{\Delta s_{\rho'_B}}{\Delta c}$.

$\mathcal{E}$ aborts in case Δc doesn't divide: Δs_e and Δs_r (abort $2a$), Δs_{r_a} and $\Delta s_{r'_a}$ (abort $2b$), Δs_{ρ_B} and $\Delta s_{\rho'_B}$ (abort $2c$). And finally, $\mathcal{E}$ aborts if Δc doesn't divide Δs_b and Δs_{ρ_B} (abort $2d$).

Therefore, after these aborts didn't happen we can infer the equivalent equalities on the right of Eqs. 2, 3, 6 and 1.

If we replace Eqs. 3 and 6 in Eq. 5 we get $1 = \left(\pm G^{\hat{r_a}} H^{\hat{r'_a}}\right)^{\Delta s_e} H^{\Delta s_\beta} G^{\Delta s_\beta} \left(\pm G^{\hat{\rho_B}} H^{\hat{\rho'_B}}\right)^{\Delta c}$ or $1 = (\pm 1)^{\Delta s_e}(\pm 1)^{\Delta c} G^{\hat{r_a}\Delta s_e + \hat{\rho_B}\Delta c + \Delta s_\beta} H^{\hat{r'_a}\Delta s_e + \hat{\rho'_B}\Delta c + \Delta s_\beta}$. Since $G, H, 1$ are quadratic residues then $(\pm 1)^{\Delta s_e}(\pm 1)^{\Delta c} = 1$, hence $1 = G^{\hat{r_a}\Delta s_e + \hat{\rho_B}\Delta c + \Delta s_\beta} H^{\hat{r'_a}\Delta s_e + \hat{\rho'_B}\Delta c + \Delta s_\beta}$. Then under the DLOG assumption $\hat{r_a}\Delta s_e + \hat{\rho_B}\Delta c + \Delta s_\beta = 0 = \hat{r'_a}\Delta s_e + \hat{\rho'_B}\Delta c + \Delta s_\beta$, which gives us that

$$\Delta s_\beta = -\hat{r_a}\Delta s_e - \hat{\rho_B}\Delta c. \tag{7}$$

Finally, we replace Eqs. 1 and 7 in Eq. 4 we get $1 = C_a^{\Delta s_e} H^{-\hat{r_a}\Delta s_e - \hat{\rho_B}\Delta c} G^{-\Delta c} \left(\pm \mathsf{Acc}^{\hat{b}} H^{\hat{\rho_B}}\right)^{\Delta c}$ or $1 = (\pm 1)^{\Delta c} C_a^{\Delta s_e} \mathsf{Acc}^{\hat{b}\Delta c} G^{-\Delta c} H^{-\hat{r_a}\Delta s_e}$ or $\left(\pm \mathsf{Acc}^{\hat{b}} G^{-1}\right)^{\Delta c} = \left(C_a^{-1} H^{r_a}\right)^{\Delta s_e}$. But as noted above Δc divides Δs_e so $\pm \mathsf{Acc}^{\hat{b}} G^{-1} = \pm \left(C_a^{-1} H^{r_a}\right)^{\hat{e}} \Rightarrow \mathsf{Acc}^{\hat{b}} G^{-1} = \pm \left(C_a^{-1} H^{\hat{r_a}}\right)^{\hat{e}} \Rightarrow \left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = \pm G$. We discern two cases:

- $\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = +G$: Then $\mathcal{E}$ sets $\tilde{D} \leftarrow \frac{C_a}{H^{\hat{r_a}}}$, $\tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$, $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ and $\tilde{b} \leftarrow \hat{b} := \frac{\Delta s_b}{\Delta c}$.
- $\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = -G$: Then $\hat{e}$ should be odd otherwise if $\hat{e} = 2\rho$ then $G = -\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{2\rho} \mathsf{Acc}^{\hat{b}}$ would be a non-quadratic residue. So $\mathcal{E}$ sets $\tilde{D} \leftarrow -\frac{C_a}{H^{\hat{r_a}}}, \tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$, $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ and $\tilde{b} \leftarrow \hat{b} := \frac{\Delta s_b}{\Delta c}$. It is clear that $\tilde{D}^{\tilde{e}} \mathsf{Acc}^{\tilde{b}} = G$.

Finally the $\mathcal{E}$ outputs $(\tilde{e}, \tilde{r}, \tilde{D}, \tilde{b})$.

Now we show that the probability the extractor terminates with outputting a valid witness is $O(\epsilon)$. If the extractor does not abort then it clearly outputs a valid witness (under the factoring assumption). For the first abort, with a standard argument it can be shown that the extractor is able to extract two accepting transcripts with probability $O(\epsilon)$ (for the probabilistic analysis we refer to [31]). Thus $Pr[\mathsf{abort1}] = 1 - O(\epsilon)$. For the aborts $\mathsf{abort}\,2a$, $\mathsf{abort}\,2b$ and $\mathsf{abort}\,2c$ they happen with negligible probability ($\leq \frac{2}{1-2^{-\lambda_s+1}} Pr[\mathcal{B}$ solves $sRSA]$ each, for any PPT adversary $\mathcal{B}$) under the strong RSA assumption according to Lemma 4.2. For $\mathsf{abort}\,2d$ we cannot directly use the same lemma as Acc is not a random generator that is part of the crs. However, with a similar argument and using partial extractability we show below that the probability for this abort is the same. Putting them together the probability of success of $\mathcal{E}$ is at least $O(\epsilon) - \frac{8}{1-2^{-\lambda_s+1}} Pr[\mathcal{B}$ solves $sRSA] = O(\epsilon) - \mathsf{negl}(\lambda_s)$.

For Eq. 1, we get from partial opening that $\mathsf{Acc} = G^{\mathsf{prod}_P}$, where $P := \{\mathsf{H}_{\mathsf{prime}}(u) \mid u \in U\}$, so

$$C_B^{\Delta c} = G^{\prod_{u \in U} \mathsf{H}_{\mathsf{prime}}(u) \cdot \Delta s_b} H^{\Delta s_{\rho_B}}.$$

We use a similar to [31] argument to prove that Δc divides Δs_b and Δs_{ρ_B} under the strong RSA assumption, given that $\lambda_s + 1 < \mu$. Then

$$C_B = \pm \mathsf{Acc}^{\hat{b}} H^{\hat{\rho_B}}. \tag{8}$$

Lemma 5.1 *Let $\lambda_s + 1 < \mu$ and $\lambda_s < \log(N)/2$ then Δc divides Δs_b and Δs_{ρ_B} under the strong RSA assumption.*

Proof An adversary against the strong RSA assumption receives $H \in \mathsf{QR}_N$ and does the following: sets $G = H^{\tau}$ for $\tau \leftarrow\$ [0, 2^{\lambda_s} N^2]$ and sends (G, H) to the adversary $\mathcal{A}$ which outputs a proof π_{Coprime}. Then we rewind to get another successful proof π'_{Coprime} and we use the extractor as above to get $C_B^{\Delta c} = G^{\prod_{u\in U} \mathsf{H_{prime}}(u)\cdot \Delta s_b} H^{\Delta s_{\rho_B}}$ or

$$C_B^{\Delta c} = H^{\tau \prod_{u\in U} \mathsf{H_{prime}}(u)\cdot \Delta s_b + \Delta s_{\rho_B}}.$$

We can exclude the case that Δc divides $\prod_{u\in U} \mathsf{H_{prime}}(u)$, since Δc is smaller than the domain of the hash function $\mathsf{H_{prime}}$, i.e. $\Delta c < \mathsf{H_{prime}}(u)$ for each $u \in U$, which comes from $\lambda_s + 1 < \mu$. Assume that $\Delta c \nmid \Delta s_b \vee \Delta c \nmid \Delta s_{\rho_B}$. we discern two cases:

- Δc doesn't divide $\tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}$: then $\gcd(\Delta c, \tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}) = g$ and there are χ, ψ such that $\chi \cdot \Delta c + \psi \cdot \left(\tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}\right) = g$. Thus

$$H^g = H^{\chi\cdot\Delta c + \psi\cdot\left(\tau \prod_{u\in U} \mathsf{H_{prime}}(u)\cdot\Delta s_b + \Delta s_{\rho_B}\right)} = H^{\chi\Delta c} \cdot C_B^{\psi\Delta c} = \left(H^{\chi} \cdot C_B^{\psi}\right)^{\Delta c}.$$

  Since g divides Δc we get $H = \pm\left(H^{\chi} \cdot C_B^{\psi}\right)^{\frac{\Delta c}{g}}$. However H is a quadratic residue (thus C_B is so), meaning that $H = \left(H^{\chi} \cdot C_B^{\psi}\right)^{\frac{\Delta c}{g}}$, thus $(H^{\chi} \cdot C_B^{\psi}, \frac{\Delta c}{g})$ is a solution to the strong RSA problem.
- Δc divides $\tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}$: let q^{ℓ} be the maximal q-power that divides Δc (i.e. q^{ℓ} is a factor of Δ) and doesn't divide at least one of Δs_b and Δs_{ρ_B}, where q is prime. Such a q^{ℓ} should exist otherwise Δc would divide both Δs_b and Δs_{ρ_B}, which we assumed it doesn't. Notice that if q^{ℓ} divided Δs_b then it would also divide Δs_{ρ_B}, as q^{ℓ} divides $\tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}$ (from assumption), so $q^{\ell} \nmid \Delta s_b$.

$$q^{\ell} \mid \left(\tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B}\right) \Rightarrow \tau \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B} = 0 \pmod{q^{\ell}}.$$

  We can write $\tau := \tau_1 + \tau_2 \,\mathsf{ord}(H)$. Notice that τ_2 is information theoretically hidden to the adversary and thus is uniformly random in $[0, 2^{\lambda_s} N^2/\mathsf{ord}(H)] \supset [0, 2^{\lambda_s} N]$ in its view.

$$\Rightarrow \tau_1 \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \tau_2 \mathsf{ord}(H) \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b + \Delta s_{\rho_B} = 0 \pmod{q^{\ell}}$$

$$\Rightarrow \tau_2 \cdot \Delta s_b = \left(-\tau_1 \prod_{u\in U} \mathsf{H_{prime}}(u) \cdot \Delta s_b - \Delta s_{\rho_B}\right) \cdot \left(\prod_{u\in U} \mathsf{H_{prime}}(u)\right)^{-1} \cdot (\mathsf{ord}(H))^{-1} \pmod{q^{\ell}}.$$

  To see that $\prod_{u\in U} \mathsf{H_{prime}}(u)$ has an inverse modulo q^{ℓ} note that since $\Delta c < \mathsf{H_{prime}}(u)$ implies $q^{\ell} < \mathsf{H_{prime}}(u)$, so $\gcd(\prod_{u\in U} \mathsf{H_{prime}}(u), q^{\ell}) = 1$. For the inverse of $\mathsf{ord}(H)$ note that $H \in \mathsf{QR}_N$ so $\mathsf{ord}(H) \in \{q_1, q_2, q_1 q_2\}$, where $N = (2q_1+1)(2q_2+1)$ is the RSA modulus. Then from $\lambda_s < \log(N)/2$ we get $\Delta c < q_1, q_2$ and thus $\gcd(\mathsf{ord}(H), q^{\ell}) = 1$. As noted above, τ_2 is uniformly random in a superset of $[0, 2^{\lambda_s} N]$. But $q^{\ell} < \Delta c < N$, so $2^{\lambda_s} N$ is at least 2^{λ_s} larger than q^{ℓ}. Thus τ_2 is statistically close to uniform in $\{0, 1, \ldots, q^{\ell} - 1\}$ (with $2^{-\lambda_s}$ error), $Pr_{\tau_2}[\tau_2 = C \pmod{q^{\ell}}] \approx \frac{1}{q^{\ell}}$. Furthermore, for

any Δs_b, $Pr_{\tau_2}[\tau_2 \cdot \Delta s_b = C \pmod{q^\ell}] \approx \frac{1}{q^\ell} \cdot \gcd(q^\ell, \Delta s_b) \leq \frac{1}{q^\ell} \cdot q^{\ell-1}$ (since q^ℓ doesn't divide Δs_b). This is because for variable τ_2, the equation $\tau_2 \Delta s_b = C \pmod{q^\ell}$ has $\gcd(q^\ell, \Delta s_b)$ solutions.
In conclusion, the probability that the above equation holds is at most $\frac{1}{q} + 2^{-\lambda_s} \leq \frac{1}{2} + 2^{-\lambda_s}$.

To summarize we showed that the probability to fall in the second case is at most $\frac{1}{2} + 2^{-\lambda_s}$. So with probability to fall in the first case, and thus solve the strong RSA problem, is at least $\frac{1}{2} - 2^{-\lambda_s}$. □

By a simple argument identical to the one of section 4.5, we can also conclude about the range of the extracted $\tilde{e}$: $s_e \overset{?}{\in} \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right]$ implies $-2^{\lambda_z+\lambda_s+\mu+2} \leq \hat{e} \leq 2^{\lambda_z+\lambda_s+\mu+2}$. □

6 A CP-SNARK for set membership in bilinear groups

In this section we propose another CP-SNARK, called $\mathsf{MemCP_{VC}}$, for the set membership relation that works in bilinear groups. Unlike the schemes of Sect. 4, the CP-SNARK given in this section does not have short parameters; specifically it has a CRS linear in the size of the sets to be committed. On the other hand, it enjoys other features that are not satisfied by our previous schemes (nor by other schemes in the literature): first, it works solely in Bilinear Groups without having to deal with RSA groups; second, it allows to commit the set in an hiding manner and, for the sake of soundness, does not need to be opened by the adversary. This is possible thanks to the fact that the set is committed in a way that (under a knowledge assumption) guarantees that the prover knows the set.

More in detail, $\mathsf{MemCP_{VC}}$ is a CP-SNARK for set membership where set elements are elements from the large field $\mathbb{F} = \mathbb{Z}_q$ where q is the order of bilinear groups. So $\mathcal{D}_{\mathsf{elm}} = \mathbb{F}$. In terms of set it supports all the subsets of $2^{\mathcal{D}_{\mathsf{elm}}}$ of cardinality bounded by n, $\mathcal{D}_{\mathsf{set}} = \{U \in 2^{\mathcal{D}_{\mathsf{elm}}} : \#U \leq n\}$, which we denote by $\mathcal{S}_n$, # symbol denotes the cardinality of a set. So U has elements in $\mathbb{F}$ and is a subset of $\mathcal{S}_n$.

6.1 Preliminaries and building blocks

6.1.1 Bilinear groups

A *bilinear group generator* $\mathcal{BG}(1^\lambda)$ outputs $(q, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e)$, where $\mathbb{G}_1$, $\mathbb{G}_2$, $\mathbb{G}_T$ are additive groups of prime order q, and $e : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is an efficiently computable, non-degenerate, bilinear map. For ease of exposition we present our results with Type-1 groups where we assume that $\mathbb{G}_1 = \mathbb{G}_2$. Our results are under the $(\ell + 1)d$-Strong Diffie Hellman and the (d, ℓ)-Extended Power Knowledge of Exponent assumptions, for which we refer the reader to [77].

6.1.2 A polynomial-pedersen type-based commitment scheme

First we present PolyCom, a type-based commitment scheme which was introduced in [18] extracted from the verifiable polynomial delegation scheme of [77]. The scheme has two types: one for ℓ-variate polynomials $f : \mathbb{F}^\ell \to \mathbb{F}$ over $\mathbb{F}$ of variable degree at most d, and one which is a standard Pedersen commitment for field elements. Let $\mathcal{W}_{\ell,d}$ be the set of all

$\mathsf{Setup}(1^\lambda, \ell, d)$: samples a bilinear group of order q, $\mathsf{bp} := (q, g, \mathbb{G}_1, \mathbb{G}_T, e) \leftarrow \mathsf{BilGen}(1^\lambda)$, samples $\alpha, \beta, s_1, \ldots, s_\ell \leftarrow \mathbb{F}$. Computes $\mathsf{prk} \leftarrow \left\{g^{\Pi_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d}\right\}$ and $\mathsf{prk}^\alpha \leftarrow \left\{g^{\alpha\cdot\Pi_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d}\right\}$. Finally samples an $s_{\ell+1} \leftarrow\$ \mathbb{F}$ and computes $h \leftarrow g^{s_{\ell+1}}$ and h^α.
Return $\mathsf{ck} \leftarrow (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^\alpha, g^\alpha, g^\beta, h, h^\alpha, h^\beta)$

$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, f)$: parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^\alpha, g^\alpha, g^\beta, h, h^\alpha, h^\beta)$ and uses $\mathsf{prk} := \left\{g^{\Pi_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d}\right\}$ and $\mathsf{prk}^\alpha := \left\{g^{\alpha\cdot\Pi_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d}\right\}$ to compute $g^{f(\boldsymbol{s})}$ and $g^{\alpha\cdot f(\boldsymbol{s})}$ respectively. Then samples a random $r_f \leftarrow\$ \mathbb{F}$ and computes $c_{f,1} \leftarrow g^{f(\boldsymbol{s})}h^{r_f}$ and $c_{f,2} \leftarrow g^{\alpha\cdot f(\boldsymbol{s})}(h^\alpha)^{r_f}$
Return $(c, o) \leftarrow ((c_{f,1}, c_{f,2}), r_f)$

$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_q, y)$: parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^\alpha, g^\alpha, g^\beta, h, h^\alpha, h^\beta)$ and samples $r \leftarrow\$ \mathbb{F}$. Computes $c_{y,1} \leftarrow g^y h^r$ and $c_{y,2} \leftarrow (g^\beta)^y(h^\beta)^r$ and return $(c, o) := ((c_{y,1}, c_{y,2}), r)$.

$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, c, f, o)$: parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^\alpha, g^\alpha, g^\beta, h, h^\alpha, h^\beta)$ and uses $\mathsf{prk} := \left\{g^{\Pi_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d}\right\}$ to compute $g^{f(\boldsymbol{s})}$. Parses $c := (c_{f,1}, c_{f,2})$. Output 1 iff $c_{f,1} = g^{f(\boldsymbol{s})}h^o \wedge e(c_{f,1}, g^\alpha) = e(c_{f,2}, g)$.

$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c, y, o)$: parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^\alpha, g^\alpha, g^\beta, h, h^\alpha, h^\beta)$. Parses $c := (c_{y,1}, c_{y,2})$ Output 1 iff $c_{y,1} = g^y h^r \wedge e(c_{y,1}, g^\beta) = e(c_{y,2}, g)$.

Fig. 13 PolyCom commitment scheme

multisets of $\{1, \ldots, \ell\}$ where the cardinality of each element is at most d. The scheme is described in Fig. 13.

Theorem 6.1 *Under the* $(\ell + 1)d$*-Strong Diffie Hellman and the* (d, ℓ)*-Extended Power Knowledge of Exponent assumptions* PolyCom *is an extractable trapdoor commitment scheme.*

For the proof we refer to [18, 77].

6.1.3 Input-hiding CP-SNARK for polynomial evaluation

The main building block of our main protocol is a CP-SNARK $\mathsf{CP}_{\mathsf{PolyEval}}$ for the type-based commitment PolyCom. Loosely speaking the idea is to commit to the input $\boldsymbol{t}$ and the output y of a polynomial (with a Pedersen commitment), further commit to the polynomial f itself (with a polynomial commitment) and then prove that the opening of the committed polynomial evaluated on the opening of the committed input gives the committed output. The relation of the protocol is $R_{\mathsf{PolyEval}}((t_k)_{k\in[\ell]}, f, y)) = 1$ iff $f(t_1, \ldots, t_\ell) = y$:

$\mathbf{R} = (\mathsf{ck}, R_{\mathsf{PolyEval}})$ where $\mathbf{R}$ is over

$$(\boldsymbol{x}, \boldsymbol{w}) = ((x, ,), (u, o, \omega)) = \big((\varnothing, (, _y , (, _{t_k})_{k\in[\ell]}, , _f)) , ((y, (t_k)_{k\in[\ell]}, f), (r_y, (r_{t_k})_{k\in[\ell]}, r_f), \varnothing) \big).$$

We will present a CP-SNARK for this relation, $\mathsf{CP}_{\mathsf{PolyEval}}$, in Sect. 6.3. $\mathsf{CP}_{\mathsf{PolyEval}}$ is based on a similar protocol for polynomial evaluation given in [18] which was in turn based on the verifiable polynomial delegation scheme of zk-vSQL [77]. In those protocols, however, the input $\mathbf{t}$ is public whereas in ours we can keep it private and committed.

6.1.4 Range proof CP-NIZK

We make use of $\mathsf{CP}_{\mathsf{Range}}$, a CP-NIZK for the following relation on PedCom commitments , and two given integers $A < B$:

$$R_{\mathsf{Range}}\left((, _e , A, B), (e, r_q)\right) = 1 \ \text{ iff } \ , = g^e h^{r_q} \ \wedge \ A < e_q < B.$$

$\mathsf{CP}_{\mathsf{Range}}$ can have various instantiations such as Bulletproofs [13].

$\mathsf{Setup}(1^\lambda, \ell)$: executes $\mathsf{ck} \leftarrow \mathsf{PolyCom.Setup}(1^\lambda, \ell, 1)$
$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_U, U)$: computes $\vec{U} \leftarrow L(U)$ and then the corresponding multilinear extension of $\vec{U}$, $f_{\vec{U}}$. Returns $(c, o) \leftarrow \mathsf{PolyCom.Commit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, f_{\vec{S}})$.
$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_q, y)$: returns $(c, o) \leftarrow \mathsf{PolyCom.Commit}(\mathsf{ck}, \mathsf{t}_q, y)$
$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, c, U, o)$: computes $\vec{U} \leftarrow L(U)$ and then the corresponding multilinear extension of $\vec{U}$, $f_{\vec{U}}$. Outputs $\mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, c, f_{\vec{S}}, o)$.
$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, c, y, o)$: outputs $\mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_q, c, y, o)$.

Fig. 14 $\mathsf{C}_{EdraxPed}$

6.1.5 Multilinear extensions of vectors

Let $\mathbb{F}$ be a field and $n = 2^\ell$. The multilinear extension of a vector $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$ in $\mathbb{F}$ is a polynomial $f_{\boldsymbol{a}} : \mathbb{F}^\ell \to \mathbb{F}$ with variables $x_1, \ldots, x_\ell$ defined as

$$f_{\boldsymbol{a}}(x_1, \ldots, x_\ell) = \sum_{i=0}^{n-1} a_i \cdot \prod_{k=1}^{\ell} \mathsf{select}_{i_k}(x_k),$$

where $i_\ell i_{\ell-1} \ldots i_2 i_1$ is the bit representation of i and $\mathsf{select}_{i_k}(x_k) = \begin{cases} x_k, & \text{if } i_k = 1 \\ 1 - x_k, & \text{if } i_k = 0. \end{cases}$

A property of Multilinear extension of $\boldsymbol{a}$ is that $f_{\boldsymbol{a}}(i_1, \ldots, i_\ell) = a_i$ for each $i \in [n]$.

6.1.6 The type-based commitment scheme of $\mathsf{MemCP}_{\mathsf{VC}}$

We define the type-based commitment $\mathsf{C}_{EdraxPed}$ for our CP-SNARK $\mathsf{MemCP}_{\mathsf{VC}}$. We recall we need a commitment that allows one to commit to both elements and sets. We build this based on a hiding variant of EDRAX Vector Commitment [23], which in turn relies on a polynomial commitment. Therefore, we use a special case of PolyCom for polynomials of maximum variable degree $d = 1$. Let $\ell := \left\lceil \log(n) \right\rceil$ and $2^{[\ell]}$ be the powerset of $[\ell] = \{1, ..., \ell\}$ then $\mathcal{W}_{\ell,1} = 2^{[\ell]}$. Furthermore, for any $n' \leq n$ let $L : \mathcal{S}_{n'} \to \mathbb{F}^{n'}$ be a function that maps a set of cardinality n' to its corresponding vector according to an ordering. The description of the scheme can be found in Fig. 14. Essentially the idea is to take the set, fix some ordering so that we can encode it with a vector, and then commit to such vector using the vector commitment of [23], which in turn commits to a vector by committing to its multilinear extension polynomial.

6.2 CP-SNARK for set membership using EDRAX vector commitment

Here we present a CP-SNARK for set membership that uses a Vector Commitment—an EDRAX [23] variant—to commit to a set. The idea is to transform a set to a vector (using for example lexicographical order) and then commit to the vector with a vector commitment. Then the set membership is proven with a zero knowledge proof of opening of the corresponding position of the vector. However to preserve zero knowledge we additionally need to hide the position of the element. For this we construct a zero knowledge proof of knowledge of an opening of a position that does not give out the position. Finally, since the position is hidden we additionally need to ensure that the prover is not cheating by providing a proof for a position that exceeds the length of the vector. For this we, also, need a proof of range for the position, i.e. that $i < n$.

CP-SNARK for R_{VCmem}

$\mathsf{KeyGen}(\mathsf{ck}, R_{\mathsf{VCmem}})$: computes $(\mathsf{ek}_1, \mathsf{vk}_1) \leftarrow \mathsf{CP}_{\mathsf{PolyEval}}.\mathsf{KeyGen}(\mathsf{ck}, R_{\mathsf{PolyEval}})$ and $(\mathsf{ek}_2, \mathsf{vk}_2) \leftarrow \mathsf{CP}_{\mathsf{Range}}.\mathsf{KeyGen}(\mathsf{ck}, R_{\mathsf{Range}})$
Return $(\mathsf{ek}, \mathsf{vk}) \leftarrow ((\mathsf{ek}_1, \mathsf{ek}_2), (\mathsf{vk}_1, \mathsf{vk}_2))$

$\mathsf{Prove}(\mathsf{ek}, \#U, (c_y, (c_{i_k})_{k\in[\ell]}, c_U), (y, (i_k)_{k\in[\ell]}, U), (r_y, (r_{i_k})_{k\in[\ell]}, r_U), \varnothing)$: parses $\mathsf{ek} := (\mathsf{ek}_1, \mathsf{ek}_2)$ and computes $\pi_1 \leftarrow \mathsf{CP}_{\mathsf{PolyEval}}.\mathsf{Prove}(\mathsf{ek}_1, \varnothing, (c_y, (c_{i_k})_{k\in[\ell]}), c_U), (y, (i_k)_{k\in[\ell]}, U), (r_y, (r_{i_k})_{k\in[\ell]}, r_U), \varnothing)$. Parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^{\alpha}, g^{\alpha}, g^{\beta}, h, h^{\alpha}, h^{\beta})$ and further $\mathsf{bp} := (q, g, \mathbb{G}_1, \mathbb{G}_T, e)$ to get (g, h), then computes $i \leftarrow \sum_{k=1}^{\ell} i_k 2^{k-1}$ and $r_i \leftarrow \sum_{k=1}^{\ell} r_{i_k} 2^{k-1}$ and the corresponding commitment $c_i \leftarrow g^i h^{r_i}$. Notice that c_i is a commitment to i with $o = r_i$. Computes $\pi_2 \leftarrow \mathsf{CP}_{\mathsf{Range}}.\mathsf{Prove}(\mathsf{ek}_1, (1, \#U), c_i, i, r_i, \varnothing)$
Return $\pi = (\pi_1, \pi_2)$

$\mathsf{VerProof}(\mathsf{vk}, \#U, (c_y, (c_{i_k})_{k\in[\ell]}, c_U), \pi)$: parses $\mathsf{vk} := (\mathsf{vk}_1, \mathsf{vk}_2)$ and $\pi := (\pi_1, \pi_2)$. Then computes homomorhically $c_{i,1} \leftarrow \prod_{k=1}^{\ell} (c_{i_k,1})^{2^{k-1}}$ and $c_{i,2} \leftarrow \prod_{k=1}^{\ell} (c_{i_k,2})^{2^{k-1}}$.
Return $b \leftarrow \mathsf{CP}_{\mathsf{PolyEval}}.\mathsf{VerProof}(\mathsf{vk}_1, \varnothing, (c_y, (c_{i_k})_{k\in[\ell]}), c_U), \pi_1) \wedge \mathsf{CP}_{\mathsf{Range}}.\mathsf{VerProof}(\mathsf{vk}_2, (1, \#U), c_i, \varnothing)$

Fig. 15 $\mathsf{MemCP}_{\mathsf{VC}}$

In this section the domain of the elements is a field, $\mathcal{D}_{\mathsf{elm}} := \mathbb{F}$, and the domain of the set is all the subsets of $2^{\mathcal{D}_{\mathsf{elm}}}$ of cardinality bounded by n, $\mathcal{D}_{\mathsf{set}} = \{U \in 2^{\mathcal{D}_{\mathsf{elm}}} : \#U \leq n\}$, which we denote by $\mathcal{S}_n$ (the # symbol denotes the cardinality of a set). So U has elements in $\mathbb{F}$ and is a subset of $\mathcal{S}_n$.

The type-based commitment of our scheme is $\mathsf{C}_{EdraxPed}$ (Fig. 14) that is presented in the previous section, and the relation is

$\mathbf{R} = (\mathsf{ck}, R_{\mathsf{VCmem}})$ where $\mathbf{R}$ is over

$$(\boldsymbol{x}, \boldsymbol{w}) = ((x, c), (u, o, \omega)) = \Big(\big(\#U, (,_y, (,_{i_k})_{k\in[\ell]}, ,_U)\big),\ \big((y, (i_k)_{k\in[\ell]}, U), (r_y, (r_{i_k})_{k\in[\ell]}, r_U), \varnothing\big)\Big).$$

$R_{\mathsf{VCmem}}(\#U, \big(y, (i_k)_{k\in[\ell]}, U\big)) = 1$ iff $y = L(U)[i] \wedge i < \#U \wedge i = \sum_{k=1}^{\ell} i_k 2^{k-1}$.

Note that in the above the prover should normally give exactly $\ell = \left\lceil \log(\#U) \right\rceil$ commitments. In case $\ell < \left\lceil \log(\#U) \right\rceil$ the position is not fully hiding since it is implicit that $i < 2^{\ell-1}$ so the verifier gets a partial information about the position.

For this we will compose a CP-SNARK $\mathsf{CP}_{\mathsf{PolyEval}}$ and a CP-NIZK $\mathsf{CP}_{\mathsf{Range}}$ for the relations $R_{\mathsf{PolyEval}}((i_k)_{k\in[\ell]}, f, y)) = 1$ iff $f(i_1, \ldots, i_\ell) = y$ and $R_{\mathsf{Range}}(T, (i_k)_{k\in[\ell]}) = 1$ iff $i < T$ respectively and the commitment scheme $\mathsf{C}_{EdraxPed}$. So $\mathsf{CP}_{\mathsf{VCmem}}$ is a conjuction of the former, where the common commitments are $(,_{i_k})_{k\in[\ell]}$ (Fig. 15).

Theorem 6.2 *Let $\mathsf{CP}_{\mathsf{PolyEval}}$ and $\mathsf{CP}_{\mathsf{Range}}$ be zero knowledge CP-SNARKs for the relations R_{PolyEval} and R_{Range} respectively under the commitment scheme* PolyCom *then the above scheme is a zero knowledge CP-SNARK for the relation R_{VCmem} and the commitment scheme $\mathsf{C}_{EdraxPed}$. Further it is a CP-SNARK for R_{mem} under the same commitment scheme.*

Proof Zero Knowledge comes directly from the zero knowledge of $\mathsf{CP}_{\mathsf{PolyEval}}$ and $\mathsf{CP}_{\mathsf{PolyEval}}$.

For Knowledge Soundness, let an adversary $\mathcal{A}(\mathbf{R}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$ outputting $(x, ,) := \big(\#U, (,_y, (,_{i_k})_{k\in[\ell]}, ,_U)\big)$ and π such that $\mathsf{VerProof}(\mathsf{vk}, \#U, (,_y, (,_{i_k})_{k\in[\ell]}, ,_U), \pi) = 1$. We will construct an extractor $\mathcal{E}$ that on input $(\mathbf{R}, \mathsf{crs}, \mathsf{aux}_R, \mathsf{aux}_Z)$ outputs a valid witness $w := \big((y, (i_k)_{k\in[\ell]}, U), (r_y, (r_{i_k})_{k\in[\ell]}, r_U), \varnothing\big)$.

$\mathcal{E}$ uses the extractors of $\mathcal{E}_{\mathsf{PolyEval}}$, $\mathcal{E}_{\mathsf{Range}}$ of $\mathsf{CP}_{\mathsf{PolyEval}}$ and $\mathsf{CP}_{\mathsf{Range}}$. $\mathcal{E}_{\mathsf{PolyEval}}$ outputs $(y, (i_k)_{k\in[\ell]}, f)$, $(r_y, (r_{i_k})_{k\in[\ell]}, r_f)$ such that $f(i_1, \ldots, i_\ell) = y \wedge \mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, ,_U , f, r_f) = 1 \wedge$ $\mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_y , y, r_y) = 1 \bigwedge_{k=1}^{\ell} \mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_{i_k}, i_k, r_{i_k}) = 1$. Further, from the Extended Power Knowledge of Exponent assumption we know that f is an ℓ-variate polynomial of maximum variable degree 1. Therefore it corresponds to a multilinear extension of a unique vector $\vec{U}$, which is efficiently computable. The extractor computes the vector $\vec{U}$ from f and the corresponding set U. It is clear that, since f is the multilinear extension of the U and $\mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_{\mathbb{F}[s]}, ,_U , f, r_f) = 1$, $\mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, ,_U , U, r_f) = 1$. $\mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_y , y, r_y) = 1 \bigwedge_{k=1}^{\ell} \mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_{i_k}, i_k, r_{i_k}) = 1$ is straightforward from the definition of the $\mathsf{C}_{Edrax\,Ped}$ commitment scheme for field elements type.

$\mathcal{E}$ uses the extractor of the commitment scheme $\mathsf{PolyCom}$, $\mathcal{E}_{\mathsf{PolyCom}}$, that outputs for each $k = 1, \ldots, \ell$ i_k, r_{i_k} such that $c_{i_k,1} = g^{i_k} h^{r_{i_k}} \wedge e(c_{i_k,1}, g^{\beta}) = e(c_{i_k,2}, g)$ or $\mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_{i_k}, r_{i_k}) = 1$. $\mathcal{E}_{\mathsf{Range}}$ outputs (i, r_i) such that $i < \#U \wedge \mathsf{PolyCom.VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_i , i, r_i) = 1$ which means that $,_{i,1} = g^i h^{r_i}$. Since the proof π is verified then $,_{i,1} = \prod_{k=1}^{\ell} (,_{i_k,1})^{2^{k-1}}$ or $g^i h^{r_i} = g^{\sum_{k=1}^{\ell} i_k 2^{k-1}} h^{\sum_{k=1}^{\ell} r_{i_k} 2^{k-1}}$. From the binding property of the Pedersen commitment we get that $i = \sum_{k=1}^{\ell} i_k 2^{k-1}$ and $r_i = \sum_{k=1}^{\ell} r_{i_k} 2^{k-1}$.

Putting them together the extractor outputs $\big((y, (i_k)_{k\in[\ell]}, U), (r_y, (r_{i_k})_{k\in[\ell]}, r_f), \varnothing\big)$ such that $\mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_y , r_y) = 1 \bigwedge_{i=1}^{\ell} \mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_q, ,_{i_k}, r_{i_k}) = 1 \wedge \mathsf{C}_{Edrax\,Ped}.\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, ,_f , U, r_f) = 1$ and further $y = L(U)[i] \wedge i < \#U \wedge i = \sum_{k=1}^{\ell} i_k 2^{k-1}$. It is straightforward that $y = L(U)[i] \wedge i < \#U$ means that $y \in U$ which leads to $R_{\mathsf{mem}}(y, U) = 1$. □

6.3 Input-hiding CP-SNARKs for polynomial evaluation

Here, we present an instantiation of a a zero knowledge CP-SNARK for the relation R_{PolyEval} presented in Sect. 6.1.

To give an intuition of the protocol we recall that zk-vSQL uses Lemma 6.1 to prove the correct evaluation of the polynomial, that we recall below.

Lemma 6.1 ([59]) *Let $f : \mathbb{F}^{\ell} \to \mathbb{F}$ be a polynomial of variable degree d. For all $\boldsymbol{t} := (t_1, \ldots, t_\ell) \in \mathbb{F}^{\ell}$ there exist efficiently computable polynomials $q_1, \ldots, q_\ell$ such that: $f(\boldsymbol{z}) - f(\boldsymbol{t}) = \sum_{i=1}^{\ell} (z_i - t_i) q_i(\boldsymbol{z})$.*

With this one can verify in time linear in the number of variables that $f(\boldsymbol{t}) = y$ by checking iff $g^{f(\boldsymbol{t})} g^{-y} = \prod_{i=1}^{\ell} e(g^{s_i}, w_i)$, given the values $g^{f(\boldsymbol{s})}$, $\{g^{s_i}\}_{i=1}^{\ell}$, $\{w_i = g^{q_i(\boldsymbol{s})}\}_{i=1}^{\ell}$ We are interested in the committed values of f, $y = f(\boldsymbol{t})$ and $\boldsymbol{t}$, $,_f$, $,_y$, $,_t$ respectively, that hide them. For this we will use instead the equation below for verification:

$$
\begin{aligned}
&\big(f(\boldsymbol{z}) + r_f z_{\ell+1}\big) - \big(f(\boldsymbol{t}) + r_y z_{\ell+1}\big) \\
&= \sum_{k=1}^{\ell} (z_k - t_k) q_k(\boldsymbol{z}) + z_{\ell+1}(r_f - r_y) \\
&= \sum_{k=1}^{\ell} (z_k - t_k)(q_k(\boldsymbol{z}) + r_k z_{\ell+1}) + z_{\ell+1} \left(r_f - r_y - \sum_{k=1}^{\ell} r_k (z_k - t_k) \right)
\end{aligned}
$$

CP-SNARK for R_{PolyEval}

$\mathsf{KeyGen}(\mathsf{ck}, R_{\mathsf{PolyEval}})$: parses $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^{\alpha}, g^{\alpha}, g^{\beta}, h, h^{\alpha}, h^{\beta})$ and computes $\mathsf{vrk} \leftarrow \{g^{s_1}, \ldots, g^{s_\ell}\}$
Return $(\mathsf{ek}, \mathsf{vk}) \leftarrow \left((\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^{\alpha}, g^{\alpha}, g^{\beta}, h, h^{\alpha}, h^{\beta}) , (\mathsf{bp}, \mathsf{vrk}, g^{\alpha}, g^{\beta}, h) \right)$

$\mathsf{Prove}(\mathsf{ek}, \varnothing, (c_y, (c_{t_k})_{k\in[\ell]}, c_f), (y, (t_k)_{k\in[\ell]}, f), (r_y, (r_{t_k})_{k\in[\ell]}, r_f), \varnothing)$: let $\mathsf{ck} := (\mathsf{bp}, \mathsf{prk}, \mathsf{prk}^{\alpha}, g^{\alpha}, g^{\beta}, h, h^{\alpha}, h^{\beta}) :=$
$\left((q, g, \mathbb{G}_1, \mathbb{G}_T, e), \left\{ g^{\prod_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d} \right\}, \left\{ g^{\alpha \cdot \prod_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d} \right\}, g^{\alpha}, g^{\beta}, g^{s_{\ell+1}}, g^{\alpha s_{\ell+1}}, g^{\beta s_{\ell+1}} \right)$ and

1. Sample $r_1, \ldots, r_\ell \leftarrow \mathbb{F}$ and compute $q_1, \ldots, q_\ell$ such that

$$(f(\boldsymbol{z}) + r_f z_{\ell+1}) - (f(\boldsymbol{t}) + r_y z_{\ell+1}) = \sum_{k=1}^{\ell} [z_k - (t_k + r_{t_k} z_{\ell+1})] \cdot [q_k(\boldsymbol{z}) + r_k z_{\ell+1}] + z_{\ell+1} \left(r_f - r_y - \sum_{k=1}^{\ell} r_k (z_k - t_k) + \sum_{k=1}^{\ell} r_{t_k} [q_k(\boldsymbol{z}) + r_k z_{\ell+1}] \right)$$

   By using $\mathsf{prk} := \left\{ g^{\prod_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d} \right\}$ and h compute $w_k = g^{q_k(\boldsymbol{s}) + r_k s_{\ell+1}}$ for each $k = 1, \ldots, \ell$ and $w_{\ell+1} = g^{r_f - r_y - \sum_{k=1}^{\ell} r_k (s_k - t_k) + \sum_{k=1}^{\ell} r_{t_k} [q_k(\boldsymbol{s}) + r_k s_{\ell+1}]}$
2. By using $\mathsf{prk}^{\alpha} := \left\{ g^{\alpha \cdot \prod_{i\in W} s_i} : W \in \mathcal{W}_{\ell,d} \right\}$ and h^{α} compute $w'_k = g^{\alpha \cdot (q_k(\boldsymbol{s}) + r_k s_{\ell+1})}$ for each $k = 1, \ldots, \ell$ and $w'_{\ell+1} = g^{\alpha \cdot (r_f - r_y - \sum_{k=1}^{\ell} r_k (s_k - t_k) + \sum_{k=1}^{\ell} r_{t_k} [q_k(\boldsymbol{s}) + r_k s_{\ell+1}])}$

Return $\pi = \{w_1, \ldots, w_\ell, w_{\ell+1}, w'_1, \ldots, w'_\ell, w'_{\ell+1}\}$

$\mathsf{VerProof}(\mathsf{vk}, \varnothing, (c_y, (c_{t_k})_{k\in[\ell]}, c_f), \pi)$: parse $\pi := \{w_1, \ldots, w_\ell, w_{\ell+1}, w'_1, \ldots, w'_\ell, w'_{\ell+1}\}$, $\mathsf{vk} := (\mathsf{bp}, \mathsf{vrk}, g^{\alpha}, g^{\beta}, h)$ and $c_y := (c_{y,1}, c_{y,2})$, $c_{t_k} := (c_{t_k,1}, c_{t_k,2})$ for each $k = 1, \ldots, \ell]$ and $c_f := (c_{f,1}, c_{f,2})$
Return 1 iff

1. $e(c_{y,1}, g^{\beta}) = e(c_{y,2}, g)$
2. $e(c_{f,1}, g^{\alpha}) = e(c_{f,2}, g)$
3. $e(c_{t_k,1}, g^{\beta}) = e(c_{t_k,2}, g)$ for all $k = 1, \ldots, \ell$
4. $e(w_k, g^{\alpha}) = e(w'_k, g)$ for all $k = 1, \ldots, \ell, \ell+1$
5. $e\left(c_f \cdot c_y^{-1}, g\right) = \prod_{k=1}^{\ell} e\left(g^{s_k} c_{t_k}^{-1}, w_k\right) \cdot e\left(g^{s_{\ell+1}}, w_{\ell+1}\right)$

Fig. 16 Description of the CP-SNARK for polynomial evaluation

$$= \sum_{k=1}^{\ell} \left[z_k - (t_k + r_{t_k} z_{\ell+1}) \right] \cdot [q_k(\boldsymbol{z}) + r_k z_{\ell+1}] + \\ + z_{\ell+1} \left(r_f - r_y - \sum_{k=1}^{\ell} r_k (z_k - t_k) + \sum_{k=1}^{\ell} r_{t_k} \left[q_k(\boldsymbol{z}) + r_k z_{\ell+1} \right] \right).$$

The equation indicates us how to construct the protocol which we present in Fig. 16.

Theorem 6.3 *Under the* $(\ell + 1)d$*-Strong Diffie Hellmann and the* (d, ℓ)*-extended power knowledge of exponent assumptions,* $\mathsf{CP}_{\mathsf{PolyEval}}$ *is a Knowledge Extractable CP-SNARK for the relation* R_{PolyEval} *and the commitment scheme* PolyCom.

Proof Below is a proof sketch, which however is quite similar to the one of $\mathsf{CP}_{\mathsf{poly}}$ in [18].

Knowledge soundness The proof comes directly from Evaluation Extractability of vSQL (see [77]) with the difference that here t_k for each $k \in [\ell]$ should also be extracted. However, its extraction is straightforward from the extractability of the commitment scheme.

Zero-knowledge Consider the following proof simulator algorithm $\mathcal{S}_{\mathsf{prv}}(\mathsf{td}, , _f , (, _{t_k})_{k\in[\ell]}, , _y)$:

- Use td to get α and $s_{\ell+1}$.
- For $k = 1$ to ℓ, sample $\xi_k \leftarrow\$ \mathbb{Z}_q$ and sets $w_k \leftarrow g^{\xi_k}$.
- Compute $w_{\ell+1}$ such that $e\left(, _{f,1} \cdot , _{y,1}^{-1}, g\right) = \prod_{k=1}^{\ell} e\left(g^{s_k}, _{t_k,1}{}^{-1}, w_k\right) \cdot e\left(g^{s_{\ell+1}}, w_{\ell+1}\right)$ holds. That is: $w_{\ell+1} \leftarrow \left(, _f \cdot c_y^{-1} \cdot \prod_{k=1}^{\ell} \left(g^{-s_k}, _{t_k,1}\right)^{\xi_k}\right)^{s_{\ell+1}^{-1}}$.
- Use α to compute $w'_k = w_k^a$ for all $k \in [\ell + 1]$.

– Return $\{w_1, ..., w_\ell, w_{\ell+1}, w'_1, \ldots, w'_\ell, w'_{\ell+1}\}$

It is straightforward to check that proofs created by $\mathcal{S}_{\mathsf{prv}}$ are identically distributed to the ones returned by $\mathsf{CP}_{\mathsf{PolyEval}}.\mathsf{Prove}$. $(w_k)_{k\in[\ell]}$'s are uniformely distributed in both cases. For $w_{\ell+1}$ there is a function W such that $w_{\ell+1} = W(,_{f,1}, ,_{y,1}, \mathsf{vk}, (,_{t_k,1})_{k\in[\ell]}, (w_k)_{k\in[\ell]})$ in both cases. Since the inputs are either identical or identically distributed, the outputs $w_{\ell+1}$ are also identically distributed in the case of of $\mathcal{S}_{\mathsf{prv}}$ and $\mathsf{CP}_{\mathsf{PolyEval}}.\mathsf{Prove}$. □

7 Experimental evaluation

We implemented all our RSA-based CP-SNARKs for set-membership and non-membership as a Rust library *cpsnarks-set*[28]. Our library is implemented in a modular fashion such that any elliptic curve from *libzexe*[67] and Ristretto from *curve25519-dalek*[54] can be used. In particular, this means that our CP-SNARKs can be easily (and efficiently) used in combination with other CP-SNARKs implemented over these elliptic curves, such as Bulletproofs [13] and LegoGroth16[19] [18].

In this section, we provide details on the implementation, we present experimental results to validate the concrete efficiency of our solutions and we compare with existing approaches.

7.1 Implementation of *cpsnarks-set*

Our *cpsnarks-set* library includes implementations of the schemes $\mathsf{MemCP}_{\mathsf{RSA}}$, $\mathsf{MemCP}_{\mathsf{RSAPrm}}$, $\mathsf{NonMemCP}_{\mathsf{RSA}}$, and $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$. In all the schemes, the RSA accumulator implementation is a modification of *accumulator*[15], and the internal protocols are implemented as interactive and are made non-interactive with the use of *Merlin*[33]. For $\mathsf{MemCP}_{\mathsf{RSA}}$ and $\mathsf{NonMemCP}_{\mathsf{RSA}}$—where we recall set elements can be binary strings and the protocol encodes them into primes—we used our implementation of LegoGroth16 [66] on top of *libzexe* to provide efficient instantiations of $\mathsf{CP}_{\mathsf{HashEq}}$. For $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ and $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$—where set elements are already primes and one needs to verify a claim about ranges—we implemented two instantiations of $\mathsf{CP}_{\mathsf{Range}}$: one based on LegoGroth16 and one based on Bulletproofs.

Each of the protocols Root, Coprime, modEq, HashEq and the different instantiations of Range are implemented individually and are further composed into the higher level membership and non-membership protocols. The higher level protocols are modular: they can use any hash-to-prime proof—or range proof in the prime elements case—as long as it implements the appropriate interface.

We benchmark the implementation on a desktop machine having a 3.8 Ghz 6-Core Intel Core i7 processor and 32GB RAM. The benchmarks code is available on [27, 28].

7.2 CP-SNARKs for set membership

For the problem of set membership, we tested the following instantiations of our solutions using the RSA-2048 [65] modulus: 1. $\mathsf{MemCP}_{\mathsf{RSA}}$ with LegoGroth16 for $\mathsf{CP}_{\mathsf{HashEq}}$ and a Blake2s-based hash-to-prime mapping to 252-bit primes ($\mathsf{MemCP}^{\mathsf{LG}}_{\mathsf{RSA}}$); 2. $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ with LegoGroth16 on the BLS12-381 curve for $\mathsf{CP}_{\mathsf{Range}}$ ($\mathsf{MemCP}^{\mathsf{LG}}_{\mathsf{RSAPrm}}$), and: (a) 252-bit

[19] We implemented this scheme in Rust on top of *libzexe* as part of this work [66].

Table 1 Set membership asymptotic complexity and benchmarks—our RSA schemes ($|x|$: size of set elements)

Solution	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$
$\text{MemCP}^{\text{LG}}_{\text{RSA}}$	$O(\|x\| \log \|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$
$\text{MemCP}^{\text{LG}}_{\text{RSAPrm}}$	$O(\|x\| \log \|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$
$\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\lambda)$	$O(\|x\| + \lambda)$

Solution	$\|x\|$	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$	P_{memory}
$\text{MemCP}^{\text{LG}}_{\text{RSA}}$	252	309.10	31.44	6852	4.4	45
$\text{MemCP}^{\text{LG}}_{\text{RSAPrm}}$	252	48.14	29.10	86	4.4	5
$\text{MemCP}^{\text{LG}}_{\text{RSAPrm}}$	63	43.91	27.492	86	4.4	5
$\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$	250	62.69	25.46	8	5.0	5
$\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$	62	38.04	21.97	2	5.0	5
	bits	ms	ms	KB	KB	MB

All the metrics of our protocols are independent of the number of elements in the set

Table 2 Set membership asymptotic complexity and benchmarks—Merkle trees through [46] zkSNARK (n: number of elements in the set)

Depth	Hash	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$
$\log n$	Pedersen	$O(\lambda \log \lambda \log n \log \log n)$	$O(\lambda)$	$O(\lambda \log n)$	$O(\lambda)$
$\log n$	SHA256	$O(\lambda \log \lambda \log n \log \log n)$	$O(\lambda)$	$O(\lambda \log n)$	$O(\lambda)$

Depth	Hash	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$	P_{memory}
8	Pedersen	216	2.8	2512	0.192	22
16	Pedersen	356	2.8	5023	0.192	35
32	Pedersen	607	2.8	10047	0.192	49
64	Pedersen	1135	2.8	20094	0.192	79
8	SHA256	1333	2.8	41276	0.192	93
16	SHA256	2563	2.8	82430	0.192	196
32	SHA256	5066	2.8	164737	0.192	423
64	SHA256	10005	2.8	329352	0.192	913
		ms	ms	KB	KB	MB

primes, (b) 63-bit primes; 3. $\text{MemCP}_{\text{RSAPrm}}$ with Bulletproofs on the Ristretto curve for CP_{Range} ($\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$), and: (a) 250-bit primes; (b) 62-bit primes.

The results of our experiments are summarized in Fig. 1.

7.2.1 Comparison with Merkle-tree approach

We compare our solutions against one based on proving a valid opening of a Merkle Tree in a SNARK. Specifically, we ran experiments for Merkle trees with maximum capacities of $\{2^8, 2^{16}, 2^{32}, 2^{64}\}$ elements, using the Groth16 SNARK [46] over the BLS12-381 curve, with the following hash functions: 1. Pedersen Hash over the Jubjub curve, a curve defined

over the scalar field of the BLS12-381 $\mathbb{G}_1$ group.[20] 2. SHA256. The Merkle tree benchmark code is based on the production Zcash code from [76]. The results of the experiments are in Fig. 2. We recall that proofs in this solution are of 192 bytes.

As one can see from the results, our solutions are highly attractive in terms of proving time and CRS size. For instance, compared to an optimized solution based on a Pedersen-Hash-based Merkle tree containing up to 2^{32} elements, our $\mathsf{MemCP}_{\mathsf{RSA}}$ scheme for arbitrary elements enjoys a sub-second proof generation on a commodity laptop, it is more than twice faster and requires a shorter CRS. A price to pay in our solution is a larger proof size (4.4 kilobytes vs. 192 bytes) and higher verification time (31 ms vs. 2.8 ms). Nevertheless, these values stay within practical reach. When comparing to less optimized solutions based on Merkle trees (e.g., using SHA256, something common in lack of specialized elliptic curves), we achieve up to 32× faster proving time and a 48× shorter CRS.

In addition to the aforementioned gains in prover efficiency, our solutions can benefit from the use of RSA accumulators to succinctly represent sets in comparison to using Merkle trees. In particular, the algebraic properties of RSA accumulators yield simple and efficient methods to add (resp. delete) elements to (resp. from) the set.

For instance, we can insert an element in an RSA accumulator in $O(1)$ time and space, and with the same complexity we can update each existing membership and non-membership witness. This means that, once having an updated witness, our zero-knowledge proofs can also be recomputed in $O(1)$ time and space. With respect to deleting elements, this can also be done in constant time and space by a party who holds a valid membership witness.

Insertion and deletion in ordinary Merkle Trees may require $O(n)$ time by rebuilding the tree from scratch from the whole set (thus also requiring $O(n)$ storage). A more efficient method for insertion requires clients to store a "frontier" of size $\Theta(\mathsf{log}(n))$ of internal hashes which lowers the time complexity to $O(\mathsf{log}(n))$. One can also lower deletion times to $O(\mathsf{log}(n))$ by using other techniques, e.g., [63], but at the expense of keeping $O(n)$ storage. Updating a Sparse Merkle Trees requires $O(n)$ time and space during updates. Inserting and deleting elements in Interval Merkle trees requires keeping the elements contiguous and sorted. This brings the time/storage complexity to $O(n)$ for insertion and deletion, since we may need to rebuild substantial portions of the tree from scratch.

7.3 CP-SNARKs for set non-membership

For set non-membership, we tested the following instantiations of our solutions using the RSA-2048 [65] modulus: 1. $\mathsf{NonMemCP}_{\mathsf{RSA}}$ with LegoGroth16 for $\mathsf{CP}_{\mathsf{HashEq}}$ and a Blake2s-based hash-to-prime mapping yielding primes of 252 bits; 2. $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$ with LegoGroth16 on the BLS12-381 curve for $\mathsf{CP}_{\mathsf{Range}}$, and 252-bit primes; 3. $\mathsf{NonMemCP}_{\mathsf{RSAPrm}}$ with Bulletproofs on the Ristretto curve for $\mathsf{CP}_{\mathsf{Range}}$, and 250-bit primes.

The results of our experiments are summarized in Fig. 3.

7.3.1 Comparison to other approaches for non-membership

Non-membership proofs are usually a more computationally intensive task in SNARKs. There are two common approaches to deal with this problem using Merkle trees: *sparse Merkle trees* and *interval Merkle trees*. We did not test these solutions experimentally. However, as we detail below, creating a zero-knowledge proof for one of these solutions would not be more

[20] This is the Bowe-Hopwood variant of a Pedersen hash, as described in [49].

Table 3 Set non-membership benchmarks—our RSA schemes ($|x|$: size of set elements)

Solution	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$
$\text{NMem}^{\text{LG}}_{\text{RSA}}$	$O(\|x\| \log \|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$
$\text{NMem}^{\text{LG}}_{\text{RSAPrm}}$	$O(\|x\| \log \|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$
$\text{NMem}^{\text{BP}}_{\text{RSAPrm}}$	$O(\|x\| + \lambda)$	$O(\|x\| + \lambda)$	$O(\lambda)$	$O(\|x\| + \lambda)$

Solution	$\|x\|$	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$	P_{memory}
$\text{NMem}^{\text{LG}}_{\text{RSA}}$	252	324.90	40.37	6852	6.1	45
$\text{NMem}^{\text{LG}}_{\text{RSAPrm}}$	252	63.39	38.12	86	6.1	5
$\text{NMem}^{\text{BP}}_{\text{RSAPrm}}$	250	79.46	34.58	8	6.6	5
	bits	ms	ms	KB	KB	MB

efficient than proving one Merkle tree path. Therefore, our solutions for non-membership achieve at least the same improvement as in the previous section.

Sparse Merkle trees for a set S are built through an ordinary Merkle Tree T on the *universe* $\mathbb{U}$ of elements (we assume there is some conventional way to index the elements). For each element x not in the set S we store a dummy element in T corresponding to the index of x. For each element in the S we store that particular element at the corresponding index. In order to prove that $x \notin S$ we provide an opening path of a Merkle tree whose leaf is a dummy value at the right index. Although there are efficient techniques to build or update a sparse Merkle Tree [4, 30], the main drawback with this technique is the opening size, which is $\Theta(\log(|\mathbb{U}|))$ instead of $\Theta(\log(|S|))$. If we perform the opening inside a SNARK, we have to pay a higher proving time. For example, consider if we use SHA256 to index elements in a set with a roughly 32 bit-representations. This would require a tree of size 2^{256} which typically implies at least a $256/32 = 8\times$ slowdown.

Interval Merkle trees work by sorting the leaves on each insertion and storing a pair of adjacent elements in each leaf, signifying intervals that don't contain elements in the set. The depth of an Interval Merkle Tree is the same as in an ordinary Merkle Tree. Nonetheless it has the following performance overheads: *(i) opening* requires two opening paths instead of only one (typically doubling the proving time); *(ii) insertion* requires sorting all leaves, which may be computationally demanding if the set is large.

Unlike either of the approaches above, the size of the set does not impact proving time in our constructions. Moreover, both insertions and non-membership witness updates are efficient to compute.

7.4 Improving running times: from statistical ZK to computational ZK

The schemes described in this section use statistically hiding commitments to achieve statistical zero-knowledge. We can improve our running times switching to computationally hiding commitments and thus computational zero-knowledge. This optimization has concrete benefits as it can cut running times by approximately half. Specifically, it reduces by 50%:

– *verification time* in constructions $\text{MemCP}_{\text{RSA}}$, $\text{MemCP}_{\text{RSAPrm}}$, $\text{NonMemCP}_{\text{RSA}}$ and $\text{NonMemCP}_{\text{RSAPrm}}$;

Table 4 Set membership benchmarks—our RSA schemes with the computational ZK optimization ($|x|$: size of set elements)

Solution	$\|x\|$	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$
$\text{MemCP}^{\text{LG}}_{\text{RSA}}$	252	292	17.6	6852	3
$\text{MemCP}^{\text{LG}}_{\text{RSAPrm}}$	252	26.11	15.12	86	3
$\text{MemCP}^{\text{LG}}_{\text{RSAPrm}}$	63	21.95	13.61	86	3
$\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$	250	41.53	11.21	8	3.6
$\text{MemCP}^{\text{BP}}_{\text{RSAPrm}}$	62	16.26	7.83	2	3.6
	bits	ms	ms	KB	KB

Table 5 Set non-membership benchmarks—our RSA schemes with the computational ZK optimization ($|x|$: size of set elements)

Solution	$\|x\|$	P_{time}	V_{time}	$\|\text{crs}\|$	$\|\Pi\|$
$\text{NMem}^{\text{LG}}_{\text{RSA}}$	252	301	22.41	6852	4.2
$\text{NMem}^{\text{LG}}_{\text{RSAPrm}}$	252	30.26	17.99	86	4.2
$\text{NMem}^{\text{BP}}_{\text{RSAPrm}}$	250	45.23	14.33	8	4.7
	bits	ms	ms	KB	KB

– *proving time* in constructions $\text{MemCP}_{\text{RSAPrm}}$ and $\text{NonMemCP}_{\text{RSAPrm}}$.

The results of our experiments for membership and non-membership are summarized in Figs. 4 and 5 respectively.

Here are more details about the optimization. Our protocols, as originally described, make use of the integer commitment of Damgard and Fujisaki [31] as described in Sect. 4.2. In this scheme we hide the value by uniformly sampling an integer r from a large set. Its size should be at least around the order of the group; for RSA groups, for example, this is equivalent to sampling $r \leftarrow\!\!\$\, [1, N/2]$. Performing exponentiations with such a large integer—on average $N/4$ in the RSA case—is expensive.

To overcome this problem, we propose a computationally hiding integer commitment variant of the above, in which r is picked from a smaller set; we sample it as $r \leftarrow\!\!\$\, [1, 2^{2\lambda}]$. The scheme is hiding under the assumption that $\{G^{r_1} : r_1 \leftarrow\!\!\$\, [1, N/2]\}$ and $\{G^{r_2} : r_2 \leftarrow\!\!\$\, [1, 2^{2\lambda}]\}$ are computationally indistinguishable.[21] This assumption can be justified in the generic group model. Similar assumption related to non-uniform distributions over $[1, \mathsf{ord}(\mathbb{G})]$ have been proven secure in GGM by Bartusek et al. [3]. This approach makes exponentiations by r faster on average since $N > 2^{2\lambda}$.

8 Applications

In this section, we discuss applications of our solutions for proving set (non-)membership in a succinct and modular way.

As one can note, in our solutions the set of committed elements is public and not hidden to the verifier. Nevertheless, our solutions can still capture some applications in which the "actual" data in the set is kept private. This is for example the case of anonymous cryptocurrencies like Zerocash. In this scenario, the public set of elements to be accumulated, U, is derived by creating a commitment to the underlying data, X, e.g., $u = COMM(x)$. To sup-

[21] Due to generic lower bounds on the DLOG problem [69], $[1, 2^{\lambda}]$ would not be enough.

port this setting, we can use our solutions for arbitrary elements (so supporting virtually any commitment scheme). Interestingly, though, we can also use our (more efficient) solution for sets of primes if commitments are prime numbers. This can be done by using for example the *hash-to-prime* method described in Sect. 4.2 or another method for Pedersen commitments that we explain below in the context of Zerocash.

We now discuss concrete applications for which our constructions are suitable, both for set-membership and set non-membership. In particular these are applications in which: (1) the prover time must be small; (2) the size of the state (i.e.: the accumulator value and commitments) must be small (potentially constant); (3) the verifier time should be small; and (4) the time to update the accumulator—adding or deleting an element—should be fast. As we discuss below, our RSA-based constructions are suitable candidate for settings with these constraints.

8.1 Zerocash

Zerocash [5] is a UTXO-type (Unspent Transaction Output) cryptocurrency protocol which extends Bitcoin with privacy-preserving (*shielded*) transactions. When performing a shielded transaction users need to prove they are spending an output note from a token they had previously received. Users concerned with privacy should not reveal which note they are spending, else their new transaction could be linked to the original note that contained the note commitment. This would reveal information *both to the public and the sender of the initial transaction, and hence partially reveal the transaction graph*. In order to keep transactions unlinkable, the protocol uses zkSNARKs to prove a set membership relation, namely that a note commitment is in a publicly known set of "usable" note commitments.

Zcash is a full-fledged digital currency using Zerocash as the underlying protocol. In its current deployment, Sapling [49], it employs Pedersen commitments of the notes and makes a zero-knowledge set membership proof of these commitments using a Pedersen-Hash-based Merkle tree approach. This is the part of the protocol that can be replaced by one of our RSA-based solutions in order to obtain a speedup in proving time. In particular, we could slightly modify the note commitments in order to enable the use of our scheme $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ for sets of prime numbers, which gives the best efficiency. We can proceed as follows. Let us recall that the note commitments are represented by their x coordinates in the underlying elliptic curve group. We can then modify them so that the sender chooses a blinding factor such that the commitment representation of a note is a prime number, and we can add a consensus rule that enforces this check. With this change, we can achieve a solution that is significantly more efficient than that currently used in Zcash. Currently Zcash uses a Merkle Tree whose depth is 32. In this setting, *we would be able to reduce proving time of set-membership from* 1.12*s to* 54.51 *ms*, trading it for larger proof sizes. We note that in this application, the set-membership proof about $u \in S$ is accompanied by another predicate $P(u)$. In the proof statement of the Zcash protocol, proving that $P(u)$ is satisfied takes considerably less time than the membership proof, hence this is why our solution would improve the overall proving time considerably, albeit the proof having more components. Another interesting comment is that our solution significantly reduces the size of the circuit, hence the need of a succinct proof system is reduced and one may even consider instantiations with other proof systems, such as Bulletproofs, that would offer transparency at the price of larger proofs and verification time.

8.2 Asset governance

In the context of *blockchain-based asset transfers protocols*, a governance system must be established to determine who can create new assets. In many cases these assets must be publicly traceable (i.e., their total supply must be public), yet in others, where the assets can be issued privately, validators still need to verify that the assets were issued by an authorized issuer. Specifically, there may be a public set of rules, X (where a $\mathsf{rule} = (pk, [a, b])$), defining which entities (public keys) are allowed to issue which assets (defined by a range of *asset types*), forming an "issuance whitelist". When one of those issuers wants to issue a new asset, they need to prove (in zero knowledge) that their public key belongs to the issuance whitelist, which entails set membership, as well as prove that the asset type they issued is within the allowed range of asset types (as defined in the original rule). In this case, the accumulated set of rules is public to all, and this public information may also include a mapping between rules and prime numbers. Our RSA-based scheme for sets of primes (Sect. 4.4) can suit this scenario.

8.3 Anonymous broadcast

In a peer-to-peer setting, anonymous broadcast allows users in a group to broadcast a message without revealing their identity. They can only broadcast *once* on each topic. One approach described in [64] works by asking users to put down a deposit which they will lose if they try and broadcast multiple messages on the same topic. In this approach users joining a group deposit their collateral in a smart contract. Whoever has the private key used by the client for the deposit can claim the sum. The approach in [64] makes sure that the key is leaked if one broadcasts more than one message. To enforce this leakage we require that at broadcast time users *(i)* derive an encryption key K that depends on their private key and the topic, and *(ii)* compute an encryption of the private key by the newly derived K. Then the users publish both the ciphertext and a secret share of the encryption key K, and prove (in zero-knowledge) their public key is part of the group and that *(i)* and *(ii)* were performed correctly. Which specific share needs to be revealed depends on the broadcasted message, thus making it likely two different shares will be leaked for two different messages.

This way, broadcasting multiple messages on the same topic reveals the user's private key, allowing other users to remove them from the group by calling a function in the smart contract and receive part of the deposit.

A particularly interesting use case for anonymous broadcast is that in which the group is comprised of validators participating in a consensus algorithm, who would like to broadcast messages without exposing their node's identity and thus prevent targeted DoS attacks. This setting requires proofs to be computed extremely fast while verification performance requirements are less strict. Our $\mathsf{MemCP}_{\mathsf{RSAPrm}}$ can satisfy these performance requirements trading for a modest increase in proof size.

8.4 Financial identities

In the financial world, regulations establish that financial organizations must *know* who their costumers are [38]. This is called a KYC check and allows to reduce the risk of fraud. Some common practices for KYC often undermines user privacy as they involve collecting a lot of personal information on them. Zero-knowledge proofs allow for an alternative approach. In modern systems, one can expect that individuals or companies will be able to prove that they

belong to a set of accepted or legitimate identities. A privacy-preserving KYC check would then be reduced to generating a set-membership proof in zero-knowledge. Often some further information is required, e.g. the credit score of the individual. In such cases our CP-SNARK for set membership can be combined with one proving an additional predicate $P(\mathsf{id})$ on the identity in a modular fashion.

Regarding applications of non-membership proofs, we expand on the well-known concept of "blacklists", where identities (or credentials) must be shown to *not* belong to a certain set of identities (or credentials). As an example, in the context of financial identities, anti-money laundering regulations (AML) [68] require customers not to be in a list of fraudulent identities. Here one can use our non-membership construction to generate a proof that the customer does not belong to the set of money launderers (or those thought to be). Because, as in the set-membership case, a user may have to prove additional information about their identity, here we can also benefit from a modular framework. Furthermore, modularity allows us to cheaply prove both membership and non-membership (at the same time) for the same identity id together with some additional information $P(\mathsf{id})$: holding commitment , (id) one can produce the following tuple of proofs: (1) a membership proof $(\mathsf{id} \in S)$; (2) a non-membership proof $(\mathsf{id} \notin S')$; (3) a CP-SNARK proof that includes the statement to be proven on that identity $(P(\mathsf{id}))$.

We note that in some cases, a central authority, who controls the white and black lists, is trusted to ensure the integrity of the lists. This means that the identities can be added or removed from the lists, which means that our RSA-based construction is ideal given the comparatively reduced cost of updating the dynamic accumulator.

8.5 Zerocoin vulnerability

Another specific application of our RSA-based constructions is that of solving the security vulnerability of the implementation of the Zerocoin protocol [56] used in the Zcoin cryptocurrency [73]. The vulnerability in a nutshell: when proving equality of values committed under the RSA commitment and the prime-order group commitment, the equality may not hold over the integers, and hence one could easily produce collisions in the prime order group. Our work can provide different ways to solve this problem by generating a proof of equality over the integers.

Acknowledgements Research leading to these results has been partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under project PICOCRYPT (Grant Agreement No. 101001283), by research grants from Protocol Labs, and from Nomadic Labs and the Tezos Foundation, by the Spanish Government under Projects SCUM (RTI2018-102043-B-I00), CRYPTOEPIC (ERC2018-092822, EUR2019-103816), PRODIGY (TED2021-132464B-I00), and RED2018-102321-T, and by the Madrid Regional Government under Project BLOQUES (S2018/TCS-4339). The last five projects are co-funded by European Union EIE, and NextGenerationEU/PRTR funds. Most of this work was done while the first author was at QEDIT. Most of this work was done while the second author was at IMDEA Software Institute and part of the work while he was at Aarhus University.

Declarations

Conflict of interest The authors have no conflicts interests to declare that are relevant to the content of this article, besides the funding that we already state and our affiliations.

A Accumulator definitions

Below is the definition of Accumulators, following the definition of [37]. We insist on public key accumulators, meaning that after the key generation phase no party has access to the secret key.

Definition A.1 *(Accumulators)* A static (non-Universal) Accumulator with domain $\mathbb{X}$ is a tuple of 4-algorithms, $\mathsf{Acc} = (\mathsf{Gen}, \mathsf{Eval}, \mathsf{Witness}, \mathsf{VerWit})$

$\mathsf{Gen}(1^\lambda, t) \to (\mathsf{sk}, \mathsf{ek}, \mathsf{vk})$ is a (probabilistic) algorithm that takes the security parameter λ and a parameter t for the upper bound of the number of elements to be accumulated. If $t = \infty$ there is no upper bound. Returns a secret key sk, an evaluation key ek and a verification key vk.
$\mathsf{Eval}(\mathsf{ek}, \mathcal{X}) \to (acc_\mathcal{X}, \mathsf{aux})$ takes the evaluation key and a set $\mathcal{X}$ and in case $\mathcal{X} \subseteq \mathbb{X}$ outputs the accumulated value $acc_\mathcal{X}$ and some auxiliary information aux. If $\mathcal{X} \nsubseteq \mathbb{X}$ outputs $\perp$.
$\mathsf{Witness}(\mathsf{ek}, x, \mathsf{aux}) \to wit_x$ takes the evaluation key ek, the value x and the auxiliary information aux and outputs either a witness wit_x of $x \in \mathcal{X}$ or $\perp$ if $x \notin \mathcal{X}$.
$\mathsf{VerWit}(\mathsf{vk}, acc_\mathcal{X}, x, w) \to b$ takes the verification key vk, the accumulation value $acc_\mathcal{X}$, a value x and a witness w and outputs 1 if wit_x is a witness of $x \in \mathcal{X}$ and 0 otherwise.

Further, we give the definition of Dynamic Accumulators, a notion that was introduced by Camenisch and Lysyanskaya [16]. Dynamic Accumulators are Accumulators that additionally provide the ability to update the accumulated value and the witnesses when the set is updated, either on addition of a new element or on deletion.

Definition A.2 *(Dynamic accumulators)* A Dynamic Accumulator Acc with domain $\mathbb{X}$ is a static Accumulator that additionally provides three algorithms $(\mathsf{Add}, \mathsf{Delete}, \mathsf{WitUpdate})$.

$\mathsf{Add}(\mathsf{ek}, acc_\mathcal{X}, y, \mathsf{aux}) \to (acc_{\mathcal{X}'}, \mathsf{aux}')$ takes the evaluation key ek, the accumulated value $acc_\mathcal{X}$, the value to be added to the set y and the auxiliary information aux. If $y \notin \mathcal{X} \wedge y \in \mathbb{X}$ outputs the new accumulation value for $\mathcal{X}' = \mathcal{X} \cup \{y\}$, $acc_{\mathcal{X}'}$ and a new auxiliary information aux'. In case $y \in \mathcal{X}$ or $y \notin \mathbb{X}$ outputs $\perp$.
$\mathsf{Delete}(\mathsf{ek}, acc_\mathcal{X}, y, \mathsf{aux}) \to (acc_{\mathcal{X}'}, \mathsf{aux}')$ takes the evaluation key ek, the accumulated value $acc_\mathcal{X}$, the value to be deleted from the set y and the auxiliary information aux. If $y \in \mathcal{X} \wedge y \in \mathbb{X}$ outputs the new accumulation value for $\mathcal{X}' = \mathcal{X} \setminus \{y\}$, $acc_{\mathcal{X}'}$ and a new auxiliary information aux'. In case $y \notin \mathcal{X}$ or $y \notin \mathbb{X}$ outputs $\perp$.
$\mathsf{WitUpdate}(\mathsf{ek}, wit_x, y, \mathsf{aux}) \to wit'_x$ takes the evaluation key ek, a witness wit_x to be updated, the value y that was either added or deleted from $\mathcal{X}$ and the auxiliary information. In case $x \in \mathcal{X}'$ outputs the updated witness wit'_x, otherwise outputs $\perp$.

Normally, we demand that update algorithms, Add and Delete are more efficient than recomputing the accumulation value from scratch with Eval. However in the publicly updatable setting this is not always possible, while it may be possible when the party holds the secret key. Still in this work we treat public key accumulators.

Security correctness

For every $t = \mathsf{poly}(\lambda)$ and $|\mathcal{X}| \leq t$:

$$Pr\left[\begin{array}{l}(\mathsf{sk},\mathsf{ek},\mathsf{vk}) \leftarrow \mathsf{GenAcc}(1^\lambda, t);\\ \quad acc_{\mathcal{X}} \leftarrow \mathsf{EvalAcc}(\mathsf{ek},\mathcal{X}); \quad : \mathsf{VerWit}(\mathsf{vk}, acc_{\mathcal{X}}, x, w)\\ \quad\quad w \leftarrow \mathsf{Witness}(\mathsf{ek},\mathcal{X},x)\end{array}\right] = 1.$$

Soundness

A cryptographic accumulator is sound if for all $t = \mathsf{poly}(\lambda)$ and for all PPT adversaries $\mathcal{A}$ there is a negligible function $\mathsf{negl}(\cdot)$ such that:

$$Pr\left[\begin{array}{l}(\mathsf{sk},\mathsf{ek},\mathsf{vk}) \leftarrow \mathsf{GenAcc}(1^\lambda, t);\ (y^*, wit_y^*, \mathcal{X}^*) \leftarrow \mathcal{A}(1^\lambda, \mathsf{ek}, \mathsf{vk});\\ acc_{\mathcal{X}^*} \leftarrow \mathsf{EvalAcc}(\mathsf{ek},\mathcal{X}^*) : \mathsf{VerWit}(\mathsf{vk}, acc_{\mathcal{X}^*}, y^*, wit_y^*) = 1 \wedge y^* \in \mathcal{X}^*\end{array}\right] \leq \mathsf{negl}(\lambda).$$

A.1 Dynamic strong RSA accumulators

We formally define Dynamic Strong RSA Accumulators [2, 6, 16] described in Sect. 4.2. It has domain $\mathbb{X} = \mathsf{Primes}$.

$\mathsf{Gen}(1^\lambda, \infty) \to (\mathsf{sk},\mathsf{ek},\mathsf{vk})$ samples an RSA modulus $(N,(q_1,q_2)) \leftarrow \mathsf{GenSRSAmod}(1^\lambda)$ and a generator $F \leftarrow_\$ \mathbb{Z}_N^*$ and computes a quadratic residue $G \leftarrow F^2 \pmod N$. Return $(\mathsf{sk},\mathsf{ek},\mathsf{vk}) \leftarrow ((q_1,q_2),(N,G),(N,G))$
$\mathsf{Eval}(\mathsf{ek},\mathcal{X}) \to (acc_{\mathcal{X}}, \mathsf{aux})$ parses $\mathsf{ek} := (N,G)$. If $\mathcal{X} \nsubseteq \mathsf{Primes}$ return $\perp$, otherwise computes $\mathsf{prod}_{\mathcal{X}} := \prod_{x_i \in \mathcal{X}} x_i$ and Return $(acc_{\mathcal{X}}, \mathsf{aux}) \leftarrow (G^{\mathsf{prod}_{\mathcal{X}}} \mod N, \mathcal{X})$
$\mathsf{Witness}(\mathsf{ek}, x, \mathsf{aux}) \to wit_x$ parses $\mathsf{ek} := (N,G)$, $\mathcal{X} := \mathsf{aux}$ and computes $\mathsf{prod}_{\mathcal{X}\setminus\{x\}} := \prod_{x_i \in \mathcal{X}\setminus\{x\}} x_i$ Return $wit_x \leftarrow G^{\mathsf{prod}_{\mathcal{X}\setminus\{x\}}} \mod N$
$\mathsf{VerWit}(\mathsf{vk}, acc_{\mathcal{X}}, x, w) \to b$ parses $\mathsf{vk} := (N,G)$ Return $b \leftarrow (w^x = acc_{\mathcal{X}} \pmod N)$.

Security of strong RSA accumulator and batch-verification

Collision Freeness of the above Accumulator comes directly from strong RSA assumption. What is more interesting is that the same scheme allows for many memberships to be verified at the same time, what is called batch-verification. That is, given $x_1, \ldots, x_m \subseteq \mathsf{Primes}$ one can compute a batch-witness $W = G^{\mathsf{prod}_{\mathcal{X}\setminus\{x_1,\ldots,x_m\}}}$ and the verification will be $b \leftarrow (W^{x_1\cdots x_m} = acc_{\mathcal{X}})$. Again the security of the batch-verification comes from strong RSA assumption and it allows us argue that for any W, x if $W^x = acc_{\mathcal{X}} := G^{\mathsf{prod}_{\mathcal{X}}}$ then $x \in \Pi_{\mathcal{X}}$, meaning that x is a product of primes of the set $\mathcal{X}$.

B Generic CP-SNARK for set membership from accumulators with proof of knowledge

We show here that any accumulator Acc scheme together with a zero knowledge proof of knowledge that a committed value is accumulated, with a commitment scheme Com, can generically construct a CP-SNARK for set membership. Let $\mathsf{CP}_{\mathsf{AccWit}}$ be a zero knowledge

$\mathsf{Setup}((1^\lambda, t))$: computes $(\mathsf{sk}, \mathsf{ek}, \mathsf{vk}) \leftarrow \mathsf{Acc.Gen}(1^\lambda, t)$ and returns $\mathsf{ck} := (\mathsf{ek}, \mathsf{vk})$.
$\mathsf{Commit}(\mathsf{ck}, \mathsf{t}_U, U)$: parses $\mathsf{ck} := (\mathsf{ek}, \mathsf{vk})$, computes $(Acc, \mathsf{aux}) \leftarrow \mathsf{Eval}(\mathsf{ek}, U)$ and returns $(c, o) := (Acc, \varnothing)$.
$\mathsf{VerCommit}(\mathsf{ck}, \mathsf{t}_U, c, U, \varnothing)$: parses $\mathsf{ck} := (\mathsf{ek}, \mathsf{vk})$, computes $(Acc, \mathsf{aux}) \leftarrow \mathsf{Eval}(\mathsf{ek}, U)$ and return 1 iff $c = Acc$.

Fig. 17 $\mathsf{Com}_{\mathsf{Acc}}$

$\mathsf{KeyGen}(\mathsf{ck}, R^{\in})$: Generate the $\mathsf{crs}_{\mathsf{AccWit}} \leftarrow\!\$\ \mathsf{CP}_{\mathsf{AccWit}}.\mathsf{KeyGen}(R_{\mathsf{AccWit}})$
Return $\mathsf{crs} := \mathsf{crs}_{\mathsf{AccWit}}$.
$\mathsf{Prove}(\mathsf{crs}, (c_U, c_u), (U, u), (\varnothing, o))$: parse $\mathsf{ck} := ((\mathsf{ek}_{\mathsf{Acc}}, \mathsf{vk}_{\mathsf{Acc}}), \mathsf{ck}_{\mathsf{Com}})$ and compute $wit_u \leftarrow \mathsf{Acc.Witness}(\mathsf{ek}, u, U)$.
Then compute $\pi_{\mathsf{AccWit}} \leftarrow \mathsf{CP}_{\mathsf{AccWit}}.\mathsf{Prove}(\mathsf{crs}, (Acc, c_u), (wit, u, o))$
Return $\pi := \pi_{\mathsf{AccWit}}$
$\mathsf{VerProof}(\mathsf{crs}, (c_U, c_u), \pi)$: Return 1 iff $\mathsf{CP}_{\mathsf{AccWit}}.\mathsf{VerProof}(\mathsf{crs}_{\mathsf{AccWit}}, (c_e, c_U), \pi_{\mathsf{Root}}) = 1$.

Fig. 18 $\mathsf{MemCP}_{\mathsf{Acc}}$ CP-SNARK for set membership

proof for the relation $R_{\mathsf{AccWit}}((Acc, ,_u), (wit, u, o)) = 1$ iff $\mathsf{VerCommit}(\mathsf{ck}, ,_u, u, o) = 1 \wedge \mathsf{VerWit}(\mathsf{vk}, Acc, u, wit) = 1$. Consider a type commitment scheme that takes one type for sets that can be accumulated by Acc and one for elements of the domain of Com. So it is the canonical composition of $\mathsf{Com}_{\mathsf{Acc}} \bullet \mathsf{Com}$, where $\mathsf{Com}_{\mathsf{Acc}}$ is decribed in Fig. 17.

Finally the generic CP-SNARK can be seen in Fig. 18.

Theorem B.1 *Let* Com *be a computationally binding commitment scheme,* Acc *a sound Accumulator scheme and* $\mathsf{CP}_{\mathsf{AccWit}}$ *be a knowledge sound proof then* $\mathsf{MemCP}_{\mathsf{Acc}}$ *is a knowledge-sound with partial opening of the set commitments* $,_U$ *for the* R_{mem} *relation and the* $\mathsf{Com}_{\mathsf{Acc}}$ *commitment scheme. Furthermore, if* Com*,is statistically hiding commitments and* $\mathsf{CP}_{\mathsf{AccWit}}$ *is zero-knowledge, then* $\mathsf{MemCP}_{\mathsf{Acc}}$ *is zero-knowledge.*

C Vector commitments

A vector commitment (VC) [21, 52] is a primitive that allows one to commit to a vector $\boldsymbol{v}$ of length n in such a way that it can later open the commitment at any position $i \in [n]$. In terms of security, a VC should be *position binding* in the sense that it is not possible to open a commitment to two different values at the same position. Also, what makes VC an interesting primitive is *conciseness*, which requires commitment and openings to be of fixed size, independent of the vector's length. Furthermore, a vector commitment can also support updates, meaning that updates in the underlying vector allow efficient updates of the commitment and the opening proofs. We note that in this case position binding should also hold with respect to updates.

C.1 Definition

We follow the definition of a Vector Commitment Scheme and its security with respect to updates as defined in [23].

Definition C.1 A Vector Commitment Scheme is tuple of PPT algorithms, $\Pi =$ (KeyGen, Com, Prove, Ver, UpdateCom, UpdateCom):

$\mathsf{KeyGen}(1^\lambda, n) \to (\mathsf{prk}, \mathsf{vrk}, \mathsf{upk}_0, \ldots, \mathsf{upk}_{n-1})$: given the security parameter λ and the size n of the committed vector it outputs a prover key prk, a verifier key vrk and update keys $\mathsf{upk}_0, \ldots, \mathsf{upk}_{n-1}$.

$\mathsf{Com}(\mathsf{prk}, a_0, \ldots, a_{n-1}) \to \mathsf{dig}_{\boldsymbol{a}}$: given prover key prk and vector $\boldsymbol{a} = (a_0, ..., a_{n-1})$, it outputs a digest $\mathsf{dig}_{\boldsymbol{a}}$ of vector $\boldsymbol{a}$.

$\mathsf{Prove}(\mathsf{prk}, i, \boldsymbol{a}) \to (a_i, \pi_i)$: given prover key prk, a vector $\boldsymbol{a} = (a_0, ..., a_{n-1})$ and an index i, it outputs the element a_i in the i-th position of the vector and a proof π_i.

$\mathsf{Ver}(\mathsf{vrk}, \mathsf{dig}, i, a, \pi) \to b$: given the verifier key prk, a digest dig, an index i, a value a and a proof π it outputs 1 iff π is a valid proof that a is in the i-th position of the vector that is committed in dig.

$\mathsf{UpdateCom}(\mathsf{dig}, i, \delta, \mathsf{upk}_i) \to \mathsf{dig}'$: given a digest dig, an index i, an update δ and an update key of i-th position it outputs an updated digest dig' of a vector the same as before but with value $a + \delta$ (instead of a) in the i-th position.

$\mathsf{UpdateProof}(\pi, i, \delta, \mathsf{upk}_i) \to \pi'$: given a digest dig, an index i, an update δ and an update key of i-th position it outputs an updated proof that $a + \delta$ (instead of a) is in the i-th position of the vector.

Soundness

A Vector Commitment Scheme Π is sound if for all PPT adversaries $\mathcal{A}$ the below probability is $\mathsf{negl}(\lambda)$

$$\Pr\left[\begin{array}{l} \mathsf{Ver}(\mathsf{vrk}, \mathsf{dig}, i, a, \pi) = 1 \\ \wedge\, a \neq a_i \end{array} : \begin{array}{r} (n, \mathsf{state}) \leftarrow \mathcal{A} \\ (\mathsf{prk}, \mathsf{vrk}, \mathsf{upk}_0, \ldots, \mathsf{upk}_{n-1}) \leftarrow \mathsf{KeyGen}(1^\lambda, n) \\ \boldsymbol{a} \leftarrow \mathcal{A} \\ \mathsf{dig} \leftarrow \mathsf{Com}(\mathsf{prk}, \boldsymbol{a}) \\ \textbf{for} k = 1, \ldots, t = \mathsf{poly}(\lambda) \\ (j, \delta) \leftarrow \mathcal{A} \\ \mathsf{dig} \leftarrow \mathsf{UpdateCom}(\mathsf{dig}, j, \delta, \mathsf{upk}_j) \\ \textbf{endfor} \\ (i, a, \pi) \leftarrow \mathcal{A} \end{array}\right] = \mathsf{negl}(\lambda).$$

C.2 EDRAX: A vector commitment from multilinear extensions

Multilinear extension of vectors

Let $\mathbb{F}$ be a field and $n = 2^\ell$. Multilinear Extension of a vector a $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$ in $\mathbb{F}$ is a polynomial $f_{\boldsymbol{a}} : \mathbb{F}^\ell \to \mathbb{F}$ with variables $x_1, \ldots, x_\ell$

$$f_{\boldsymbol{a}}(x_1, \ldots, x_\ell) = \sum_{i=0}^{n-1} a_i \cdot \prod_{k=1}^{\ell} \mathsf{select}_{i_k}(x_k),$$

where $i_\ell i_{\ell-1} \ldots i_2 i_1$ is the bit representation of i and $\mathsf{select}_{i_k}(x_k) = \begin{cases} x_k, & \text{if } i_k = 1 \\ 1 - x_k, & \text{if } i_k = 0. \end{cases}$

A property of Multilinear extension of $\boldsymbol{a}$ is that $f_{\boldsymbol{a}}(i_1, \ldots, i_\ell) = a_i$ for each $i \in [n]$.

Vector commitment scheme

We describe the EDRAX Vector Commitment:

Definition C.2 Let a bilinear group $\mathsf{bp} = (q, g, \mathbb{G}_1, \mathbb{G}_T, e) \leftarrow \mathcal{RG}(1^\lambda)$ generated by a group generator. Let $n = 2^\ell$ be the length of the vector and $2^{[\ell]}$ be the powerset of $[\ell] = \{1, \ldots, \ell\}$

$\mathsf{KeyGen}(1^\lambda, n) \rightarrow (\mathsf{prk}, \mathsf{vrk}, \mathsf{upk}_0, \ldots, \mathsf{upk}_{n-1})$: samples random $s_1, \ldots, s_\ell \leftarrow\$ \mathbb{F}$ and computes $\mathsf{prk} \leftarrow \left\{g^{\prod_{i \in S} s_i} : S \in 2^{[\ell]}\right\}$ and $\mathsf{vrk} \leftarrow \{g^{s_1}, \ldots, g^{s_\ell}\}$. For each $i = 0, \ldots, n-1$ computes the update key $\mathsf{upk}_i \leftarrow \left\{g^{\prod_{k=1}^{t} \mathsf{select}_{i_k}(s_k)} : t = 1, \ldots, \ell\right\} := \{\mathsf{upk}_{i,t} : t = 1, \ldots, \ell\}$.

$\mathsf{Com}(\mathsf{prk}, a_0, \ldots, a_{n-1}) \rightarrow \mathsf{dig}_{\boldsymbol{a}}$: let $\boldsymbol{a} := (a_0, \ldots, a_{n-1})$. Computes $\mathsf{dig}_{\boldsymbol{a}} \leftarrow g^{f_{\boldsymbol{a}}(s_1, \ldots, s_\ell)}$ where $f_{\boldsymbol{a}}$ is the multilinear extension of vector $\boldsymbol{a}$ as described above.

$\mathsf{Prove}(\mathsf{prk}, i, \boldsymbol{a}) \rightarrow (a_i, \pi_i)$: let $\boldsymbol{x} = (x_1, \ldots, x_\ell)$ be an ℓ-variable. Compute $q_1, \ldots, q_\ell$ such that $f_{\boldsymbol{a}}(\boldsymbol{x}) - f_{\boldsymbol{a}}(i_1, \ldots, i_\ell) = \sum_{k=1}^{\ell}(x_k - i_k)q_k(\boldsymbol{x})$ and $\pi_i \leftarrow \left\{g^{q_1(\boldsymbol{s})}, \ldots, g^{q_\ell(\boldsymbol{s})}\right\}$ (where $g^{q_i(\boldsymbol{s})}$ is evaluated by using $\mathsf{prk} := \left\{g^{\prod_{i \in S} s_i} : S \in 2^{[\ell]}\right\}$ without $\boldsymbol{s}$).

$\mathsf{Ver}(\mathsf{vrk}, \mathsf{dig}, i, a, \pi) \rightarrow b$: parse $\pi := (w_1, \ldots, w_\ell)$ and outputs 1 iff $e(\mathsf{dig}/g^a, g) = \prod_{k=1}^{\ell} e(g^{s_k - i_k}, w_k)$

$\mathsf{UpdateCom}(\mathsf{dig}, i, \delta, \mathsf{upk}_i) \rightarrow \mathsf{dig}'$: computes $dig' \leftarrow dig \cdot \left[g^{\prod_{k=1}^{\ell} \mathsf{select}_{i_k}(s_k)}\right]^\delta := \mathsf{dig} \cdot \left[\mathsf{upk}_{i,\ell}\right]^\delta = g^{(a_i + \delta) \cdot \prod_{k=1}^{\ell} \mathsf{select}_{i_k}(s_k) + \sum_{j=0, j \neq i}^{n-1} a_j \cdot \prod_{k=1}^{\ell} \mathsf{select}_{j_k}(s_k)}$

$\mathsf{UpdateCom}(\pi, i, a', \mathsf{upk}_i) \rightarrow \pi'$: Parses $\pi := (w_1, \ldots, w_\ell)$ and computes $w'_k \leftarrow w_k \cdot g^{\Delta_i(\boldsymbol{s})}$ for each $k = 1, \ldots, \ell$, where $\Delta_k(\boldsymbol{x})$ are the delta polynomials computed by the DELTAPOLYNOMIALS algorithm (for more details about the algorithm and its correctness we refer to [23]).

The above scheme is proven in [23] to satisfy the Soundness property under the q-Strong Bilinear Diffie-Hellman assumption.

D Another instantiation of protocol for R_{Coprime}

Below we propose another interactive ZK protocol for R_{Coprime}. The difference with the above is that it doesn't have the limitation of $\lambda_s + 1 < \mu$ and $\lambda_s < \log(N)/2$. Also, partial opening of Acc isn't needed. This comes with a cost of 2 more group elements in the proof size, 4 more exponentiations for the prover and 2 more for the verifier.

1. Prover computes $C_a = DH^{r_a}, C_{r_a} = G^{r_a}H^{r'_a}, C_b = G^b H^{\rho_b}, C_B = Acc^b H^{\rho_B}, C_{\rho_B} = G^{\rho_B}H^{\rho'_B}$ and sends to the verifier:
   $\mathcal{P} \rightarrow \mathcal{V} : C_a, C_b, C_{r_a}, C_B, C_{\rho_B}$.
2. Prover and Verifier perform a protocol for the relation: $R((\mathsf{Acc}, C_e, C_a, C_{r_a}, C_b, C_B, C_{\rho_B}), (e, r, r_a, r'_a, b, \rho_b, \rho_B, \rho'_B, D, B, \beta, \delta)) = 1$ iff

$$C_b = G^b H^{\rho_b} \wedge C_B = Acc^b H^{\rho_B} \wedge C_e = G^e H^r \wedge C_{r_a} = G^{r_a} H^{r'_a}$$
$$\wedge C_{\rho_B} = G^{\rho_B} H^{\rho'_B} \wedge C_a^e C_B = GH^\beta \wedge C_{r_a}^e C_{\rho_B} = G^\beta H^\delta.$$

Coprime2 protocol

On common reference string $\mathsf{crs} = (\mathbb{Z}_N^*, G, H)$

$\mathsf{Prove}(\mathsf{crs}, (C_e, \mathsf{Acc}), (e, r, (D, b)))$:

1. samples $r_a, r_{a'}, \rho_b, \rho_B, \rho_{B'} \leftarrow\$ \left(-\lfloor N/4 \rfloor, \lfloor N/4 \rfloor\right)$ and computes $C_a = DH^{r_a}, C_{r_a} = G^{r_a} H^{r'_a}, C_b = G^b H^{\rho_b}, C_B = Acc^b H^{\rho_B}, C_{\rho_B} = G^{\rho_B} H^{\rho'_B}$.
2. Computes the non-interactive version of the above protocol
   $r_b, r_e \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right)$
   $r_{\rho_b}, r_{\rho_B}, r_r, r_{r_a}, r_{r'_a}, r_{\rho'_B} \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$
   $r_\beta, r_\delta \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right)$
   $\alpha_1 = G^{r_b} H^{r_{\rho_b}}, \alpha_2 = \mathsf{Acc}^{r_b} H^{r_{\rho_B}}, \alpha_3 = G^{r_e} H^{r_r}, \alpha_4 = G^{r_{ra}} H^{r_{r'_a}}, \alpha_5 = C_a^{r_e} H^{r_\beta}, \alpha_6 = C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta}, \alpha_7 = G^{r_{\rho_B}} H^{r_{\rho'_B}}$
   $c \leftarrow \mathsf{H}(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, C_e, Acc)$
   $s_b = r_b - cb, s_e = r_e - ce, s_{\rho_b} = r_{\rho_b} - c\rho_b, s_{\rho_B} = r_{\rho_B} - c\rho_B, s_r = r_r - cr, s_{r_a} = r_{r_a} - cr_a, s_{r'_a} = r_{r'_a} - cr_{r_a}, s_{\rho'_B} = r_{\rho'_B} - c\rho'_B, s_\beta = r_\beta + c(er_a + \rho_B), s_\delta = r_\delta + c(er'_a + \rho'_B)$

   Returns $\pi \leftarrow (C_a, C_{r_a}, C_b, C_B, C_{\rho_B}, \alpha_1, \alpha_2, \alpha_3, \alpha_4, cs_b, s_e, s_{\rho_b}, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta)$

$\mathsf{VerProof}(\mathsf{crs}, (C_e, Acc), \pi)$: returns 1 iff $\alpha_1 = C_b^c G^{s_b} H^{s_{\rho_b}} \wedge \alpha_2 = C_B^c \mathsf{Acc}^{s_b} H^{s_{\rho_B}} \wedge \alpha_3 = C_e^c G^{s_e} H^{s_r} \wedge \alpha_4 = C_{r_a}^c G^{s_{ra}} H^{s_{r'_a}} \wedge \alpha_5 = C_a^{s_e} H^{s_\beta} G^c C_B^{-c} \wedge \alpha_6 = C_{r_a}^{s_e} H^{s_\delta} G^{s_\beta} C_{\rho_B}^{-c} \wedge \alpha_7 = C_{\rho_B}^c G^{s_{\rho_B}} H^{s_{\rho'_B}} \wedge \wedge s_e \in \left[-2^{\lambda+\lambda_s+\mu+1}, 2^{\lambda+\lambda_s+\mu+1}\right]$

Fig. 19 Description of the alternative Coprime protocol

Let λ_s be the size of the challenge space, λ_z be the statistical security parameter and μ the size of e.

– Prover samples:

$$r_b, r_e \leftarrow\$ \left(-2^{\lambda_z+\lambda_s+\mu}, 2^{\lambda_z+\lambda_s+\mu}\right)$$
$$r_{\rho_b}, r_{\rho_B}, r_r, r_{r_a}, r_{r'_a}, r_{\rho'_B} \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s}\right)$$
$$r_\beta, r_\delta \leftarrow\$ \left(-\lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}, \lfloor N/4 \rfloor 2^{\lambda_z+\lambda_s+\mu}\right),$$

and computes:

$$\alpha_1 = G^{r_b} H^{r_{\rho_b}}, \quad \alpha_2 = \mathsf{Acc}^{r_b} H^{r_{\rho_B}}, \quad \alpha_3 = G^{r_e} H^{r_r}, \quad \alpha_4 = G^{r_{ra}} H^{r_{r'_a}},$$
$$\alpha_5 = C_a^{r_e} H^{r_\beta}, \quad \alpha_6 = C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta}, \quad \alpha_7 = G^{r_{\rho_B}} H^{r_{\rho'_B}}.$$

$\underline{\mathcal{P} \rightarrow \mathcal{V}} : (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7)$

– Verifier samples the challenge $c \leftarrow \{0,1\}^{\lambda_s}$ $\underline{\mathcal{V} \rightarrow \mathcal{P}} : c$

– Prover computes the response:

$$s_b = r_b - cb, \quad s_e = r_e - ce$$
$$s_{\rho_b} = r_{\rho_b} - c\rho_b, s_{\rho_B} = r_{\rho_B} - c\rho_B, s_r = r_r - cr, s_{r_a} = r_{r_a} - cr_a, s_{r'_a} = r_{r'_a} - cr'_a,$$
$$s_{\rho'_B} = r_{\rho'_B} - c\rho'_B$$
$$s_\beta = r_\beta + c(er_a + \rho_B), \quad s_\delta = r_\delta + c(er'_a + \rho'_B).$$

$\underline{\mathcal{P} \rightarrow \mathcal{V}} : (s_b, s_e, s_{\rho_b}, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta)$.

– Verifier checks if:

$$\alpha_1 \stackrel{?}{=} C_b^c G^{s_b} H^{s_{\rho_b}}, \quad \alpha_2 \stackrel{?}{=} C_B^c \mathsf{Acc}^{s_b} H^{s_{\rho_B}}, \quad \alpha_3 \stackrel{?}{=} C_e^c G^{s_e} H^{s_r}, \quad \alpha_4 \stackrel{?}{=} C_{r_a}^c G^{s_{ra}} H^{s_{r'_a}},$$
$$\alpha_5 \stackrel{?}{=} C_a^{s_e} H^{s_\beta} G^c C_B^{-c}, \quad \alpha_6 \stackrel{?}{=} C_{r_a}^{s_e} H^{s_\delta} G^{s_\beta} C_{\rho_B}^{-c}, \quad \alpha_7 \stackrel{?}{=} C_{\rho_B}^c G^{s_{\rho_B}} H^{s_{\rho'_B}},$$
$$s_e \stackrel{?}{\in} \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right].$$

Correctness

Here we show the correctness of the protocol (Fig. 19).

$$
\begin{aligned}
\alpha_1 &= G^{r_b} H^{r_{\rho_b}} = G^{s_b + cb} H^{s_{\rho_b} + cr_b} = G^{s_b} H^{s_{\rho_b}} (G^b H^{\rho_b})^c \\
&= G^{s_b} H^{s_{\rho_b}} C_b^c \\
\alpha_2 &= \mathsf{Acc}^{r_b} H^{r_{\rho_B}} = \mathsf{Acc}^{s_b + cb} H^{s_{\rho_B} + c\rho_B} = \mathsf{Acc}^{s_{r_b}} H^{s_{\rho_B}} (\mathsf{Acc}^b H^{\rho_B})^c \\
&= \mathsf{Acc}^{s_b} H^{s_{\rho_B}} C_B^c \\
\alpha_3 &= G^{r_e} H^{r_r} = G^{s_e + ce} H^{s_r + cr} = G^{s_e} H^{s_r} (G^e H^r)^c \\
&= G^{s_e} H^{s_r} C_e^c \\
\alpha_4 &= G^{r_{ra}} H^{r_{r'_a}} = G^{s_{ra} + cr_a} H^{s_{r'_a} + cr'_a} = G^{s_{ra}} H^{s_{r'_a}} (G^{r_a} H^{r'_a})^c \\
&= G^{s_{ra}} H^{s_{r'_a}} C_{r_a}^c \\
\alpha_5 &= C_a^{r_e} H^{r_\beta} = C_a^{s_e + ce} H^{s_\beta - c(er_a + \rho_B)} = C_a^{s_e} H^{s_\beta} (D^e H^{er_a})^c H^{-c(er_a + \rho_B)} \\
&= C_a^{s_e} H^{s_\beta} (D^e H^{-\rho_B})^c = C_a^{s_e} H^{s_\beta} (GAcc^{-b} H^{-\rho_B})^c = \\
&= C_a^{s_e} H^{s_\beta} G^c C_B^{-c} \\
\alpha_6 &= C_{r_a}^{r_e} G^{r_\beta} H^{r_\delta} = C_{r_a}^{s_e + ce} G^{s_\beta - c(er_a + \rho_B)} H^{s_\delta - c(er'_a + \rho'_B)} \\
&= C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} (G^{r_a} H^{r'_a})^{ce} G^{-c(er_a + \rho_B)} H^{-c(er'_a + \rho'_B)} = C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} G^{-c\rho_B} H^{-c\rho'_B} \\
&= C_{r_a}^{s_e} G^{s_\beta} H^{s_\delta} C_{\rho_B}^{-c} \\
\alpha_7 &= G^{r_{\rho_B}} H^{r_{\rho'_B}} = G^{s_{\rho_B} + c\rho_B} H^{s_{\rho'_B} + c\rho'_B} = G^{s_{\rho_B}} H^{s_{\rho'_B}} (G^{\rho_B} H^{\rho'_B})^c \\
&= G^{s_{\rho_B}} H^{s_{\rho'_B}} C_{\rho_B}^c .
\end{aligned}
$$

Security

Theorem D.1 *Let $\mathbb{Z}_N^*$ be an RSA group where strong-RSA assumption holds, then the above protocol is an honest-verifier zero knowledge and knowledge sound protocol for R_{Coprime}.*

Proof Zero-Knowledge can be proven with standard techniques, similar to the ones in the proof of Theorem 4.6 and is therefore omitted.

For the knowledge soundness, let an adversary of the knowledge soundness $\mathcal{A}$ that is able to convince the verifier $\mathcal{V}$ with a probability at least ϵ. We will construct an extractor $\mathcal{E}$ that extracts the witness $(e, r, r_2, r_3, \beta, \delta)$. Using rewinding $\mathcal{E}$ gets two accepted transcripts

$$
\begin{aligned}
&(C_a, C_b, C_{r_a}, C_B, C_{\rho_B}, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, c, s_b, s_e, s_{\rho_b}, s_{\rho_B}, s_r, s_{r_a}, s_{r'_a}, s_{\rho'_B}, s_\beta, s_\delta) \\
&(C_a, C_b, C_{r_a}, C_B, C_{\rho_B}, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, c', s'_b, s'_e, s'_{\rho_b}, s'_{\rho_B}, s'_r, s'_{r_a}, s'_{r'_a}, s'_{\rho'_B}, s'_\beta, s'_\delta),
\end{aligned}
$$

on two different challenges c and c'. $\mathcal{E}$ aborts if it cannot get two such transcripts (abort1).

We denote $\Delta c := c' - c$, $\Delta s_b := s_b - s'_b$, $\Delta s_e := s_e - s'_e$, $\Delta s_{\rho_b} := s_{\rho_b} - s'_{\rho_b}$, $\Delta s_{\rho_B} := s_{\rho_B} - s'_{\rho_B}$, $\Delta s_r := s_r - s'_r$, $\Delta s_{r_a} := s_{r_a} - s'_{r_a}$, $\Delta s_{r'_a} := s_{r'_a} - s'_{r'_a}$, $\Delta s_{\rho'_B} := s_{\rho'_B} - s'_{\rho'_B}$, $\Delta s_\beta := s_\beta - s'_\beta$, $\Delta s_\delta := s_\delta - s'_\delta$ then

$$
C_b^{\Delta c} = G^{\Delta s_b} H^{\Delta s_{\rho_b}} \Rightarrow C_b = \pm G^{\hat{b}} H^{\hat{\rho_b}}, \tag{9}
$$

$$C_B^{\Delta c} = \mathsf{Acc}^{\Delta s_b} H^{\Delta s_{\rho_B}} \Rightarrow C_B = \pm \mathsf{Acc}^{\hat{b}} H^{\hat{\rho_B}}, \tag{10}$$

$$C_e^{\Delta c} = G^{\Delta s_e} H^{\Delta s_r} \Rightarrow C_e = \pm G^{\hat{e}} H^{\hat{r}}, \tag{11}$$

$$C_{r_a}^{\Delta c} = G^{\Delta s_{ra}} H^{\Delta s_{r_a'}} \Rightarrow C_{ra} = \pm G^{\hat{r_a}} H^{\hat{r_a'}}, \tag{12}$$

$$1 = C_a^{\Delta s_e} H^{\Delta s_\beta} G^{-\Delta c} C_B^{\Delta c}, \tag{13}$$

$$1 = C_{r_a}^{\Delta s_e} H^{\Delta s_\delta} G^{\Delta s_\beta} C_{\rho_B}^{\Delta c}, \tag{14}$$

$$C_{\rho_B}^{\Delta c} = G^{\Delta s_{\rho_B}} H^{\Delta s_{\rho_B'}} \Rightarrow C_{\rho_B} = \pm G^{\hat{\rho_B}} H^{\hat{\rho_B'}}, \tag{15}$$

define the (possibly rational) numbers $\hat{b} := \frac{\Delta s_b}{\Delta c}$, $\hat{\rho_b} := \frac{\Delta s_{\rho_b}}{\Delta c}$, $\hat{e} := \frac{\Delta s_e}{\Delta c}$, $\hat{r} := \frac{\Delta s_r}{\Delta c}$, $\hat{r_a} := \frac{\Delta s_{ra}}{\Delta c}$, $\hat{r_a'} := \frac{\Delta s_{r_a'}}{\Delta c}$, $\hat{\rho_B} := \frac{\Delta s_{\rho_B}}{\Delta c}$, $\hat{\rho_B'} := \frac{\Delta s_{\rho_B'}}{\Delta c}$.

$\mathcal{E}$ aborts in case Δc doesn't divide: Δs_b and Δs_{ρ_b}(abort $2a$), Δs_e and Δs_r(abort $2b$), Δs_{ra} and $\Delta s_{r_a'}$(abort $2c$), Δs_{ρ_B} and $\Delta s_{\rho_B'}$(abort $2d$). And finally, $\mathcal{E}$ aborts if Δc doesn't divide Δs_{ρ_B} (abort $2e$). Therefore, after these aborts didn't happen we can infer the equivalent equalities on the right of Eqs. 9, 10, 11, 12 and 15.

If we replace Eqs. 12 and 15 in Eq. 14 we get $1 = \left(\pm G^{\hat{r_a}} H^{\hat{r_a'}}\right)^{\Delta s_e} H^{\Delta s_\beta} G^{\Delta s_\beta} \left(\pm G^{\hat{\rho_B}} H^{\hat{\rho_B'}}\right)^{\Delta c}$ or $1 = (\pm 1)^{\Delta s_e} (\pm 1)^{\Delta c} G^{\hat{r_a} \Delta s_e + \hat{\rho_B} \Delta c + \Delta s_\beta} H^{\hat{r_a'} \Delta s_e + \hat{\rho_B'} \Delta c + \Delta s_\beta}$. Since $G, H, 1$ are quadratic residues then $(\pm 1)^{\Delta s_e} (\pm 1)^{\Delta c} = 1$, hence $1 = G^{\hat{r_a} \Delta s_e + \hat{\rho_B} \Delta c + \Delta s_\beta} H^{\hat{r_a'} \Delta s_e + \hat{\rho_B'} \Delta c + \Delta s_\beta}$. Then under the DLOG assumption $\hat{r_a} \Delta s_e + \hat{\rho_B} \Delta c + \Delta s_\beta = 0 = \hat{r_a'} \Delta s_e + \hat{\rho_B'} \Delta c + \Delta s_\beta$, which gives us that

$$\Delta s_\beta = -\hat{r_a} \Delta s_e - \hat{\rho_B} \Delta c. \tag{16}$$

Finally, we replace Eqs. 10 and 16 in Eq. 13 we get $1 = C_a^{\Delta s_e} H^{-\hat{r_a} \Delta s_e - \hat{\rho_B} \Delta c} G^{-\Delta c} \left(\pm \mathsf{Acc}^{\hat{b}} H^{\hat{\rho_B}}\right)^{\Delta c}$ or $1 = (\pm 1)^{\Delta c} C_a^{\Delta s_e} \mathsf{Acc}^{\hat{b} \Delta c} G^{-\Delta c} H^{-\hat{r_a} \Delta s_e}$ or $\left(\pm \mathsf{Acc}^{\hat{b}} G^{-1}\right)^{\Delta c} = \left(C_a^{-1} H^{r_a}\right)^{\Delta s_e}$. But as noted above Δc divides Δs_e so $\pm \mathsf{Acc}^{\hat{b}} G^{-1} = \pm \left(C_a^{-1} H^{r_a}\right)^{\hat{e}} \Rightarrow \mathsf{Acc}^{\hat{b}} G^{-1} = \pm \left(C_a^{-1} H^{\hat{r_a}}\right)^{\hat{e}} \Rightarrow \left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = \pm G$. We discern two cases:

- $\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = +G$: Then $\mathcal{E}$ sets $\tilde{D} \leftarrow \frac{C_a}{H^{\hat{r_a}}}$, $\tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$, $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ and $\tilde{b} \leftarrow \hat{b} := \frac{\Delta s_b}{\Delta c}$.
- $\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{\hat{e}} \mathsf{Acc}^{\hat{b}} = -G$: Then $\hat{e}$ should be odd otherwise if $\hat{e} = 2\rho$ then $G = -\left(\frac{C_a}{H^{\hat{r_a}}}\right)^{2\rho} \mathsf{Acc}^{\hat{b}}$ would be a non-quadratic residue. So $\mathcal{E}$ sets $\tilde{D} \leftarrow -\frac{C_a}{H^{\hat{r_a}}}$, $\tilde{e} \leftarrow \hat{e} := \frac{\Delta s_e}{\Delta c}$, $\tilde{r} \leftarrow \hat{r} := \frac{\Delta s_r}{\Delta c}$ and $\tilde{b} \leftarrow \hat{b} := \frac{\Delta s_b}{\Delta c}$. It is clear that $\tilde{D}^{\tilde{e}} \mathsf{Acc}^{\tilde{b}} = G$.

Finally the $\mathcal{E}$ outputs $(\tilde{e}, \tilde{r}, \tilde{D}, \tilde{b})$.

Now we show that the probability the extractor terminates with outputting a valid witness is $O(\epsilon)$. If the extractor does not abort then it clearly outputs a valid witness (under factoring assumption). For the first abort, with a standard argument it can be shown that the extractor is able to extract two accepting transcripts with probability $O(\epsilon)$ (for the probabilistic analysis we refer to [31]). Thus $Pr[\text{abort1}] = 1 - O(\epsilon)$. For the aborts abort $2a$, abort $2b$, abort $2c$ and abort $2d$ they happen with negligible probability ($\leq \frac{2}{1-2^{-\lambda_s+1}} Pr[\mathcal{B} \text{ solves } sRSA]$ each, for any PPT adversary $\mathcal{B}$) under the strong RSA assumption according to Lemma 4.2. For abort $2e$ we show in the lemma below that in case it happens an adversary can solve the

strong RSA problem. Putting them together the probability of success of $\mathcal{E}$ is at least $O(\epsilon) - \left(\frac{8}{1-2^{-\lambda_s+1}} + 1\right) Pr[\mathcal{B} \text{ solves } sRSA] = O(\epsilon) - \mathsf{negl}(\lambda_s)$.

Lemma D.1 *If Δc divides Δs_b then it also divides $\Delta \rho_B$ under the strong RSA assumption.*

Proof An adversary to the strong RSA assumption receives $H \in \mathsf{QR}_N$ and does the following: set $G = H^\tau$ for $\tau \leftarrow\$ [0, 2^{\lambda_s} N^2]$ and send (G, H) to the adversary $\mathcal{A}$ which outputs a proof π_{Coprime2}. Then we rewind to get another successful proof π'_{Coprime2} and we use the extractor as above to get $C_B^{\Delta c} = \mathsf{Acc}^{\Delta s_b} H^{\Delta s_{\rho_B}}$.

Assume that $\Delta c \nmid \Delta \rho_B$. Since Δc divides Δs_b then there is a k such that $k \cdot \Delta c = \Delta s_b$. Then $C_B^{\Delta c} = \mathsf{Acc}^{k \cdot \Delta c} H^{\Delta s_{\rho_B}} \Rightarrow \left(C_B \mathsf{Acc}^{-k}\right)^{\Delta c} = H^{\Delta s_{\rho_B}}$. From assumption Δc doesn't divide $\Delta \rho_B$, so $\gcd(\Delta c, \Delta \rho_B) = g$ for a $g \neq \Delta c, \Delta \rho_B$. Hence, there are there are χ, ψ such that $\chi \cdot \Delta c + \psi \cdot \Delta \rho_B = g$. Thus, $H^g = H^{\chi \cdot \Delta c + \psi \cdot \Delta \rho_B} = H^{\chi \Delta c} \left(C_B \mathsf{Acc}^{-k}\right)^{\psi \Delta c} = \left(H^\chi C_B^\psi \mathsf{Acc}^{-\psi k}\right)^{\Delta c}$ so $H = \pm \left(H^\chi C_B^\psi \mathsf{Acc}^{-\psi k}\right)^{\frac{\Delta c}{g}}$. Now since H and Acc are quadratic residues (and so is C_B) we get that $H = \left(H^\chi C_B^\psi \mathsf{Acc}^{-\psi k}\right)^{\frac{\Delta c}{g}}$ and thus $\left(H^\chi C_B^\psi \mathsf{Acc}^{-\psi k}, \frac{\Delta c}{g}\right)$ is a solution to the strong RSA problem. □

By a simple argument identical to the one of Sect. 4.5, we can also conclude about the range of the extracted $\tilde{e}$: $s_e \overset{?}{\in} \left[-2^{\lambda_z+\lambda_s+\mu+1}, 2^{\lambda_z+\lambda_s+\mu+1}\right]$ implies $-2^{\lambda_z+\lambda_s+\mu+2} \leq \hat{e} \leq 2^{\lambda_z+\lambda_s+\mu+2}$. □

E Instantiation over hidden order groups

In Sects. 4 and 5 we construct zero knowledge protocols for set membership/non-membership, where the sets are committed using an RSA accumulator. The integer commitment scheme IntCom, the RSA accumulator-based commitments to sets $\mathsf{SetCom}_{\mathsf{RSA}}, \mathsf{SetCom}_{\mathsf{RSA}'}$, the proof of equality modEq, the argument of knowledge of a root Root and the argument of knowledge of coprime element $\mathsf{Coprime}$ are all working over RSA groups.

Although in our work above we specify the group to be an RSA group, we note that our protocols can also work over any Hidden Order Group. For example Class Groups [12] or the recently proposed groups from Hyperelliptic Curves [34, 51].

Here we describe the (slight) modifications, in the protocols and the assumptions under which they would be secure, that are necessary to switch to (general) Hidden Order Groups.

Let $\mathsf{Ggen}(1^\lambda)$ be a probabilistic algorithm that generates such a group $\mathbb{G}$ with order in a specific range $[\mathsf{ord}_{min}, \mathsf{ord}_{max}]$ such that $\frac{1}{\mathsf{ord}_{min}}, \frac{1}{\mathsf{ord}_{max}}, \frac{1}{\mathsf{ord}_{max} - \mathsf{ord}_{min}} \in \mathsf{negl}(\lambda)$.

The additional assumption that we need to make is that it is hard to find any group element in $\mathbb{G}$ of low (poly-size) order. This is the Low Order Assumption [10], which is formally defined below:

Definition E.1 (*Low order assumption* [10]) We say that the *low order assumption* holds for a Ggen if for any PPT adversary $\mathcal{A}$:

$$\Pr\left[\begin{array}{l} u^\ell = 1 \\ \wedge u \neq 1 \\ \wedge 1 < \ell < 2^{\mathsf{poly}(\lambda)} \end{array} : \begin{array}{l} \mathbb{G} \leftarrow \mathsf{Ggen}(\lambda) \\ (u, \ell) \leftarrow \mathcal{A}(\mathbb{G}) \end{array}\right] = \mathsf{negl}(\lambda).$$

We note that specifically for RSA groups, for Low Order assumption to hold, we have to work in the quotient group $\mathbb{Z}_N^*/\{1,-1\}$ [72], since otherwise -1 would trivially break the assumption. So $\mathbb{Z}_N^*/\{1,-1\}$ would be an instantiation of a Hidden Order Group where the Low Order assumption holds.

In terms of constructions, one difference regards the upper bound on the order of $\mathbb{G}$ that is used in the protocols. More precisely, throughout the main core of our work we use N as an upper bound for the order of the group $\mathbb{Z}_N^*$ and $N/2$ as an upper bound for the order of the quadratic residues subgroup QR_N. Similarly, in a Hidden Order Group $\mathbb{G}$ generated by Ggen, although the order of the group is unknown, a range in which the order lies is known $[\mathsf{ord}_{min}, \mathsf{ord}_{max}]$. So the maximum order ord_{max} can be used, instead of N, as an upper bound. In many cases these values are used either to securely sample a random value or to bound the size of a value needed for a security proof. For example a random value that is sampled from $\left(-\lfloor N/4\rfloor 2^{\lambda_z+\lambda_s}, \lfloor N/4\rfloor 2^{\lambda_z+\lambda_s}\right)$ in the RSA group instantiation will be sampled from $\left(-\frac{\mathsf{ord}_{max}}{2}2^{\lambda_z+\lambda_s}, \frac{\mathsf{ord}_{max}}{2}2^{\lambda_z+\lambda_s}\right)$ in the case of hidden order groups.

Here we give other specific changes that need to be made to instantiate our protocols in general hidden order groups. For IntCom, the verification equation becomes $C = G^x H^r$ (without the $\pm$). Then the argument of knowledge of opening of such a commitment would be secure under the strong RSA and low order assumptions. The set commitments $\mathsf{SetCom}_{\mathsf{RSA}}$, $\mathsf{SetCom}_{\mathsf{RSA}'}$ remain the same and are binding under the strong RSA assumption for Ggen (and collision resistance of $\mathsf{H}_{\mathsf{prime}}$ for the case of $\mathsf{SetCom}_{\mathsf{RSA}}$). For modEq, the same difference as for the AoK of an opening of an IntCom commitment is inherited. For Root and $\mathsf{Coprime}$, the Proposition 4.1 needs to be slightly modified: $A = B^{\frac{x}{y}}$ can be without $\pm$, and can be proven under the low order assumption instead. Finally, in the proof of security of protocol $\mathsf{Coprime}$, in Lemma 5.1 the assumption $\lambda_s < \log(N)/2$ is not needed as long as the low order assumption holds (an adversary that can find $H, \Delta c$ such that $\gcd(\mathsf{ord}(H), q^\ell) = 1$ can be used to break low order assumption).

Transparent instantiation and efficiency

The above instantiation combined with a transparent proof system (for instance Bulletproofs) gives transparent CP-NIZKs for set (non)-membership analogously with the ones described for RSA groups in Sect. 4, i.e. proof systems with a uniformly random CRS. We ran some preliminary experiments for this instantiation over class groups of 2048-bit discriminant and using Bulletproofs. The results showed proving time of 3.3 s, verification time of 2.3 s and proof size of 5.3 kB, for arbitrary accumulated elements (i.e. not necessarily primes). Furthermore, if we make use of the optimization described in Sect. 7.4 it boosts the efficiency to 1.66 s, 1.33 s and 4 kB (prover/verifier and proof size resp.).

Unfortunately, very recent cryptanalytic results on class groups [34] showed that a discriminant of 2048 bits yields only about 60 bits of security level, while for 128 bits of security one needs to choose a 6600-bit discriminant for the class group. We estimate that over class groups of a 6000-bit discriminant our aforementioned protocol, together with the optimization of Sect. 7.4, will give proving time of $\sim$ 12 s, verification time of $\sim$ 6.4 s and proof size of 6.4 kB. Finally, our estimations for the respective protocol for prime elements (with the computational ZK optimization) are: $\sim$ 7 s/$\sim$ 6.2 s/6 KB (proving time/verification time/proof size resp.).

References

1. Agrawal S., Ganesh C., Mohassel P.: Non-interactive zero-knowledge proofs for composite statements. In: Shacham H., Boldyreva A. (eds.) CRYPTO 2018, pp. 643–673. Part III, volume 10993 of LNCS. Springer, Heidelberg (2018)
2. Bari N., Pfitzmann B.: Collision-free accumulators and fail-stop signature schemes without trees. In: Fumy W. (ed.) EUROCRYPT'97, vol. 1233, pp. 480–494. LNCS. Springer, Heidelberg (1997).
3. Bartusek J., Ma F., Zhandry M.: The distinction between fixed and random generators in group-based assumptions. In: Shacham H., Boldyreva A. (eds.) CRYPTO 2019, pp. 801–830. Part II, LNCS. Springer, Heidelberg (2019)
4. Ben L., Emilia K.: Revocation transparency. Google Research, September, p. 33 (2012)
5. Ben-Sasson E., Chiesa A., Garman C., Green M., Miers I., Tromer E., Virza M.: Zerocash: decentralized anonymous payments from bitcoin. In: 2014 IEEE Symposium on Security and Privacy, pp. 459–474. IEEE Computer Society Press (2014)
6. Benaloh J.C., de Mare M.: One-way accumulators: a decentralized alternative to digital sinatures (extended abstract). In: Helleseth T. (ed.) EUROCRYPT'93, vol. 765, pp. 274–285. LNCS. Springer, Heidelberg (1994).
7. Benarroch D., Campanelli M., Fiore D.: Community standards proposal for commit-and-prove zero-knowledge proof systems (2019). https://www.binarywhales.com/assets/misc/zkproof-cp-standards.pdf
8. Benarroch D., Campanelli M., Fiore D., Gurkan K., Kolonelos D.: Zero-knowledge proofs for set membership: efficient, succinct, modular. In: International Conference on Financial Cryptography and Data Security, pp. 393–414. Springer (2021)
9. Benoît L., San L., Khoa N., Huaxiong W.: Zero-knowledge arguments for lattice-based accumulators: logarithmic-size ring signatures and group signatures without trapdoors. In: Marc F., Jean-Sébastien C. (eds.) EUROCRYPT 2016, pp. 1–31. Part II, volume 9666 of LNCS. Springer, Heidelberg (2016)
10. Boneh D., Bünz B., Fisch B.: A survey of two verifiable delay functions. Cryptology ePrint Archive, Report 2018/712 (2018). https://eprint.iacr.org/2018/712
11. Boneh D., Bünz B., Fisch B.: Batching techniques for accumulators with applications to iops and stateless blockchains. IACR Cryptol. ePrint Arch. **2018**, 1188 (2018).
12. Buchmann J., Hamdy S.: A survey on IQ cryptography (2001). http://tubiblio.ulb.tu-darmstadt.de/100933/
13. Bünz B., Bootle J., Boneh D., Poelstra A., Wuille P., Maxwell G.: Bulletproofs: short proofs for confidential transactions and more. In: 2018 IEEE Symposium on Security and Privacy, pages 315–334. IEEE Computer Society Press (2018)
14. Cachin C., Micali S., Stadler M.: Computationally private information retrieval with polylogarithmic communication. In: Stern J. (ed.) EUROCRYPT'99, vol. 1592, pp. 402–414. LNCS. Springer, Heidelberg (1999).
15. Cambrian Tech: Cryptographic accumulators in rust (2019). https://github.com/cambrian/accumulator
16. Camenisch J., Lysyanskaya A.: Dynamic accumulators and application to efficient revocation of anonymous credentials. In: Yung M. (ed.) CRYPTO 2002, vol. 2442, pp. 61–76. LNCS. Springer, Heidelberg (2002).
17. Camenisch J., Kohlweiss M., Soriente C.: An accumulator based on bilinear maps and efficient revocation for anonymous credentials. In: Jarecki S., Tsudik G. (eds.) PKC 2009, vol. 5443, pp. 481–500. LNCS. Springer, Heidelberg (2009).
18. Campanelli M., Fiore D., Querol A.: Legosnark: modular design and composition of succinct zero-knowledge proofs. To appear at ACM CCS 2019. IACR Cryptology ePrint Archive, 2019 (2019)
19. Campanelli M., Fiore D., Han S., Kim J., Kolonelos D., Oh H.: Succinct zero-knowledge batch proofs for set accumulators. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, pp. 455–469 (2022)
20. Canetti R., Lindell Y., Ostrovsky R., Sahai A.: Universally composable two-party and multi-party secure computation. In: 34th ACM STOC, pp. 494–503. ACM Press (2002)
21. Catalano D., Fiore D.: Vector commitments and their applications. In: Kurosawa K., Hanaoka G. (eds.) PKC 2013, volume 7778 of LNCS, pp. 55–72. Springer, Heidelberg (2013).
22. Chaum D.: Security without identification: transaction systems to make big brother obsolete. Commun. ACM **28**(10), 1030–1044 (1985).
23. Chepurnoy A., Papamanthou C., Yupeng Z.: A cryptocurrency with stateless transaction validation, Edrax (2018)
24. Couteau G., Hartmann D.: Shorter non-interactive zero-knowledge arguments and zaps for algebraic languages. In: Advances in Cryptology–CRYPTO 2020: 40th Annual International Cryptology Conference, CRYPTO 2020, Santa Barbara, CA, USA, August 17–21, 2020, Proceedings, Part III, pp. 768–798. Springer (2020)

25. Couteau G., Peters T., Pointcheval D.: Removing the strong RSA assumption from arguments over the integers. In: Coron J.-S., Nielsen J.B. (eds.) EUROCRYPT 2017, Part II, volume 10211 of LNCS, pp. 321–350. Springer, Heidelberg (2017).
26. Couteau G., Lipmaa H., Parisella R., Ødegaard A.T.: Efficient nizks for algebraic sets. In: Advances in Cryptology–ASIACRYPT 2021: 27th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 6–10, (2021), Proceedings, Part III, pp. 128–158. Springer (2021)
27. cpsnarks-librustzcash. https://github.com/kobigurk/cpsnarks-librustzcash
28. Cpsnarks-set. https://github.com/kobigurk/cpsnarks-set
29. Cramer R., Shoup V.: Signature schemes based on the strong RSA assumption. In: Motiwalla J., Tsudik G. (eds) ACM CCS 99, pp. 46–51. ACM Press (1999)
30. Dahlberg R., Pulls T., Peeters R.: Efficient sparse merkle trees: Caching strategies and secure (non-)membership proofs. Cryptology ePrint Archive, Report 2016/683 (2016). https://eprint.iacr.org/2016/683
31. Damgård I., Fujisaki E.: A statistically-hiding integer commitment scheme based on groups with hidden order. In: Yuliang Zheng (ed.) ASIACRYPT 2002, vol. 2501, pp. 125–142. LNCS. Springer, Heidelberg (2002).
32. Damgård I., Triandopoulos N.: Supporting non-membership proofs with bilinear-map accumulators. Cryptology ePrint Archive, Report 2008/538 (2008). http://eprint.iacr.org/2008/538
33. de Valence, H.: Merlin: composable proof transcripts for public-coin arguments of knowledge (2019). https://github.com/dalek-cryptography/merlin
34. Dobson S., Galbraith Steven D.: Trustless groups of unknown order with hyperelliptic curves. Cryptology ePrint Archive, Report 2020/196 (2020). https://eprint.iacr.org/2020/196
35. Eagen L., Fiore D., Gabizon A.: cq: Cached quotients for fast lookups. Cryptology ePrint Archive (2022)
36. Escala A., Groth J.: Fine-tuning Groth-Sahai proofs. In: Krawczyk H. (ed.) PKC 2014, vol. 8383, pp. 630–649. LNCS. Springer, Heidelberg (2014).
37. Fazio N., Nicolosi A.: Cryptographic accumulators: definitions, constructions and applications. Paper written for course at New York University. www.cs.nyu.edu/nicolosi/papers/accumulators.pdf (2002)
38. FINRA: https://www.finra.org/rules-guidance/rulebooks/finra-rules/2090#the-rule
39. Fiore D., Fournet C., Ghosh E., Kohlweiss M., Ohrimenko O., Parno B.: Hash first, argue later: adaptive verifiable computations on outsourced data. In: Weippl E.R., Katzenbeisser S., Kruegel C., Myers A.C., Halevi S.(eds) ACM CCS 2016, pp. 1304–1316. ACM Press (2016)
40. Fouque P.-A., Tibouchi M.: Close to uniform prime number generation with fewer random bits. In: Esparza J., Fraigniaud P., Husfeldt T., Koutsoupias E. (eds.) ICALP 2014, pp. 991–1002. Part I, volume 8572 of LNCS. Springer, Heidelberg (2014)
41. Fujisaki E., Okamoto T.: Statistical zero knowledge protocols to prove modular polynomial relations. In: Kaliski B.S. (ed.) CRYPTO'97, volume 1294 of LNCS, pp. 16–30. Springer, Heidelberg (1997).
42. Gabizon A., Khovratovich D.: Flookup: Fractional decomposition-based lookups in quasi-linear time independent of table size. Cryptology ePrint Archive (2022)
43. Gennaro R., Halevi S., Rabin T.: Secure hash-and-sign signatures without the random oracle. In: Stern J. (ed.) EUROCRYPT'99, vol. 1592, pp. 123–139. LNCS. Springer, Heidelberg (1999).
44. Gentry C., Wichs D.: Separating succinct non-interactive arguments from all falsifiable assumptions. In: Fortnow L., Vadhan S.P. (eds) 43rd ACM STOC, pp. 99–108. ACM Press (2011)
45. Goldwasser S., Micali S., Rackoff C.: The knowledge complexity of interactive proof systems. SIAM J. Comput. **18**(1), 186–208 (1989).
46. Groth J.: On the size of pairing-based non-interactive arguments. In: Fischlin M., Coron J.-S. (eds.) EUROCRYPT 2016, pp. 305–326. Part II, volume 9666 of LNCS. Springer, Heidelberg (2016)
47. Groth J., Sahai A.: Efficient non-interactive proof systems for bilinear groups. In: Smart N.P. (ed.) EUROCRYPT 2008, vol. 4965, pp. 415–432. LNCS. Springer, Heidelberg (2008).
48. Helger L.: Secure accumulators from euclidean rings without trusted setup. In: Feng B., Pierangela S., Jianying Z. (eds.) ACNS 12, vol. 7341, pp. 224–240. LNCS. Springer, Heidelberg (2012).
49. Hopwood D., Bowe S., Hornby T., Wilcox N.: Zcash protocol specification. Tech. rep. 2016–1.10. Zerocoin Electric Coin Company, Tech. Rep., (2016). https://github.com/zcash/zips/blob/master/protocol/sapling.pdf
50. Jiangtao L., Ninghui L., Rui X.: Universal accumulators with efficient nonmembership proofs. In: Jonathan K., Moti Y. (eds.) ACNS 07, vol. 4521, pp. 253–269. LNCS. Springer, Heidelberg (2007).
51. Lee J.: The security of groups of unknown order based on jacobians of hyperelliptic curves. Cryptology ePrint Archive, Report 2020/289 (2020). https://eprint.iacr.org/2020/289

52. Libert B., Yung M.: Concise mercurial vector commitments and independent zero-knowledge sets with short proofs. In: Micciancio D. (ed.) TCC 2010, vol. 5978, pp. 499–517. LNCS. Springer, Heidelberg (2010).
53. Lipmaa H., Parisella R.: Set (non-) membership nizks from determinantal accumulators. Cryptology ePrint Archive (2022)
54. Lovecruft I.A., de Valence H.: curve25519-dalek: a pure-rust implementation of group operations on ristretto and curve25519. https://github.com/dalek-cryptography/curve25519-dalek
55. Merkle R.C.: A digital signature based on a conventional encryption function. In: Pomerance C. (ed.) CRYPTO'87, vol. 293, pp. 369–378. LNCS. Springer, Heidelberg (1988).
56. Miers I., Garman C., Green M., Rubin Aviel D: Zerocoin: anonymous distributed E-cash from Bitcoin. In: 2013 IEEE Symposium on Security and Privacy, pp. 397–411. IEEE Computer Society Press (2013)
57. Nguyen L.: Accumulators from bilinear pairings and applications. In: Menezes A. (ed.) CT-RSA 2005, vol. 3376, pp. 275–292. LNCS. Springer, Heidelberg (2005).
58. Ozdemir A., Wahby Riad S., Whitehat B., Boneh D.: Scaling verifiable computation using efficient set accumulators. Cryptology ePrint Archive, Report 2019/1494 (2019). https://eprint.iacr.org/2019/1494
59. Papamanthou C., Shi E., Tamassia R.: Signatures of correct computation. In: Sahai A. (ed.) TCC 2013, vol. 7785, pp. 222–242. LNCS. Springer, Heidelberg (2013).
60. Papamanthou C., Shi E., Tamassia R., Yi K.: Streaming authenticated data structures. In: Johansson T., Nguyen P.Q. (eds.) EUROCRYPT 2013, vol. 7881, pp. 353–370. LNCS. Springer, Heidelberg (2013).
61. Parno B., Howell J., Gentry C., Raykova M.: Pinocchio: nearly practical verifiable computation. In: 2013 IEEE Symposium on Security and Privacy, pp. 238–252. IEEE Computer Society Press (2013)
62. Posen J., Kattis Assimakis A: Caulk+: table-independent lookup arguments. Cryptology ePrint Archive (2022)
63. Ray J.: Patricia tree (2019). https://github.com/ethereum/wiki/wiki/Patricia-Tree
64. rln Semaphore: rate limiting nullifier for spam prevention in anonymous p2p setting, February (2019). https://ethresear.ch/t/semaphore-rln-rate-limiting-nullifier-for-spam-prevention-in-anonymous-p2p-setting/5009
65. Rsa-2048. https://en.wikipedia.org/wiki/RSA_numbers#RSA-2048
66. Rust implementation of LegoGroth16. https://github.com/kobigurk/legogro16
67. SCIPR Lab: Zexe (zero knowledge execution). https://github.com/scipr-lab/zexe
68. Securities U.S. and Exchange Commission: Anti-money laundering (aml) source tool for broker-dealers (2018). https://www.sec.gov/about/offices/ocie/amlsourcetool.htm
69. Shoup V.: Lower bounds for discrete logarithms and related problems. In: Fumy W. (ed.) EUROCRYPT'97, vol. 1233, pp. 256–266. LNCS. Springer, Heidelberg (1997).
70. Srinivasan S., Karantaidou I., Baldimtsi F., Papamanthou C.: Batching, aggregation, and zero-knowledge proofs in bilinear accumulators. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, pp. 2719–2733 (2022)
71. Valiant P.: Incrementally verifiable computation or proofs of knowledge imply time/space efficiency. In: Canetti R. (ed.) TCC 2008, vol. 4948, pp. 1–18. LNCS. Springer, Heidelberg (2008).
72. Wesolowski B.: Efficient verifiable delay functions. Cryptology ePrint Archive, Report 2018/623 (2018). https://eprint.iacr.org/2018/623
73. Yap R.: Cryptographic description of zerocoin attack (2019). https://zcoin.io/cryptographic-description-of-zerocoin-attack/
74. Zapico A., Buterin V., Khovratovich D., Maller M., Nitulescu A., Simkin M.: Caulk: lookup arguments in sublinear time. In:Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, pp. 3121–3134 (2022)
75. Zapico A., Gabizon A., Khovratovich D., Maller M., Carla R.: Nearly optimal lookup arguments. Cryptology ePrint Archive, Baloo (2022).
76. Zcash: Zcash rust crates. https://github.com/zcash/librustzcash
77. Zhang Y., Genkin D., Katz J., Papadopoulos D., Papamanthou C.: A zero-knowledge version of vSQL. Cryptology ePrint Archive, Report 2017/1146 (2017). https://eprint.iacr.org/2017/1146
78. Zhang Y., Katz J., Papamanthou C.: An expressive (zero-knowledge) set accumulator. In: 2017 IEEE European Symposium on Security and Privacy (EuroS P), pp. 158–173 (2017)
79. Zhang J., Xie T., Zhang Y., Song D.: Transparent polynomial delegation and its applications to zero knowledge proof. In: IEEE Symposium on Security and Privacy (2020)