

# Earn While You Reveal: Private Set Intersection that Rewards Participants

**Abstract.** In Private Set Intersection protocols (PSIs), a non-empty result always reveals information about the private input sets of the parties. Moreover, in various variants of PSI, not all parties necessarily receive or are interested in the result. However, to date, the literature has assumed that those parties not receiving or not interested in the result still contribute their private input sets to the PSI for free of charge, although doing so would cost them their privacy. In this work, we introduce a novel multi-party PSI, called “Anesidora”, which *rewards* parties for contributing their private input sets to the protocol. Anesidora is designed with efficiency in mind, primarily relying on symmetric key primitives. Its complexities scale linearly with the number of parties and set cardinality. It maintains security even when the majority of parties are compromised by malicious colluding adversaries. Along the way, we devised “Justitia”, a pioneering *fair multi-party PSI* and introduced the concept of *unforgeable polynomials*, which may hold independent significance.

## 1 Introduction

Secure Multi-Party Computation (MPC) is a general-purpose cryptographic protocol that enables multiple mutually distrustful parties to collectively compute a certain functionality on their private inputs, safeguarding their privacy beyond the result. Private Set Intersection (PSI) is a sub-class of MPC, that allows parties to compute the intersection of their private sets. It aims to efficiently attain the same security as MPC.

Two facts about PSIs exist: (i) a non-empty result always reveals something about the parties’ private input sets (i.e., the set elements that are in the intersection), and (ii) various variants of PSIs do not output the result to all parties, even in those PSIs that do, not all of the parties are necessarily interested in it. Given these facts, a natural question arises: *How can we incentivise parties that do not receive the result or do not express interest in it to participate in a PSI?*

To date, the literature has not addressed the aforementioned question. It has assumed that all parties would engage in a PSI for free of charge, resulting in them bearing the privacy burden (in addition to costs associated with the PSI).<sup>1</sup>

---

<sup>1</sup> Combining a multi-party PSI with a payment mechanism that charges a fixed amount to the initiating buyer presents significant issues. Firstly, this approach compels the buyer to pay, even if malicious clients tamper with result accuracy, enabling them to gain knowledge without the buyer learning the correct result, as there is no existing fair multi-party PSI in the literature. Secondly, it enforces a fixed fee irrespective of the actual intersection size. Overcharging dissuades buyers from participation while undercharging discourages other clients from engaging in the protocol. This highlights the need for a more equitable and flexible payment model in such scenarios.

In this work, we answer the above question for the first time. We present a multi-party PSI, called “Anesidora”, enabling the buyer who initiates the PSI computation (and has an interest in the result) to pay other parties proportionally based on the number of elements it learns about other parties’ private inputs. Anesidora is efficient and predominantly *relies on symmetric key primitives*. Its computation and communication complexities scale linearly with the number of parties and set cardinality. Anesidora maintains security even in scenarios where the majority of parties are compromised by malicious adversaries who may collude.

We construct Anesidora in a modular manner. Specifically, we introduce the formal concept of “PSI with Fair Compensation” ( $\mathcal{PSI}^{\mathcal{FC}}$ ) and devise “Justitia” which realises this concept.  $\mathcal{PSI}^{\mathcal{FC}}$  guarantees that either all parties get the result, or in the event of an unfair protocol abortion (where only dishonest parties gain access to the result), honest parties will receive financial compensation. Justitia stands as *the inaugural* fair multi-party PSI.

Subsequently, we advance from the  $\mathcal{PSI}^{\mathcal{FC}}$  to the “PSI with Fair Compensation and Reward” ( $\mathcal{PSI}^{\mathcal{FCR}}$ ) notion and develop Anesidora to actualise  $\mathcal{PSI}^{\mathcal{FCR}}$ . The latter notion guarantees that honest parties (a) receive rewards irrespective of the honesty of all parties or if a subset aborts unfairly, and (b) are compensated in the case of an unfair termination. We prove the two PSIs using the simulation-based model. In pursuit of efficient PSIs, we introduce a novel primitive, called unforgeable polynomials, which may independently pique interest.

Designing Justitia and Anesidora presented particular challenges. In their development, we needed to tackle several issues, including (i) minimising overhead, (ii) safeguarding against collusion among malicious parties to prevent unfair payment, (iii) ensuring the transfer of the *intact* output from one subroutine as input to another, (iv) providing clients with rewards commensurate with the intersection size, and (v) maintaining the confidentiality of outgoing messages.

A PSI, like Anesidora, supporting more than two parties and rewarding set contributors can offer opportunities for richer analytics, and incentivise participation. It can be utilised in various scenarios: (1) by advertisers targeting specific customers by identifying shared shopping patterns across different e-commerce databases [27], (2) by a malware detection service querying multiple antivirus databases to identify applications as malware [44], or (3) by a bank like “We-Bank” [12] that uses “vertical federated learning” and PSI to gather customer data from partners to enhance loan risk management. In these instances, set contributors receive rewards through this PSI.

We hope that our work serves as a catalyst for future research in:

- developing reward mechanisms for participants in *generic* MPC. These mechanisms have the potential to enhance the real-world adoption of MPC.
- bridging between two research domains: “data market” [25,34,45] and MPC.

**Our Contributions Summary.** In this work, we: (1) introduce Anesidora, the pioneering PSI enabling participants to receive rewards for contributing their set elements to the intersection, (2) introduce Justitia, the first fair multi-party PSI, and (3) provide formal definitions of the aforementioned constructions.

## 2 Related Work

Since their introduction in [19], various PSIs have been designed, categorised into *traditional* and *delegated* types. In traditional PSIs, data owners compute the result interactively using their local data. In this research domain, Raghuraman and Rindal [42] proposed two two-party PSIs, one secure against semi-honest/passive and the other against malicious/active adversaries. To date, these protocols are the fastest two-party PSIs. They rely on Oblivious Key-Value Stores (OKVS) data structure and Vector Oblivious Linear Evaluation (VOLE). These PSIs’ computation cost is  $O(c)$ , where  $c$  is a set’s cardinality. They also impose  $O(c \log c^2 + \kappa)$  and  $O(c \cdot \kappa)$  communication costs in the semi-honest and malicious models respectively, where  $\kappa$  is a security parameter. Also, researchers designed PSIs that enable multiple (i.e., more than two) parties to efficiently compute the intersection. The multi-party PSIs in [26,35] are secure against passive adversaries while those in [8,22,47,35,38] were designed to remain secure against active ones. To date, the protocols in [35] and [38] are the most efficient multi-party PSIs designed to be secure against passive and active adversaries respectively. They maintain security even if the majority of parties are corrupt. The overall computation and communication complexities of the PSI in [35] are  $O(c \cdot m^2 + c \cdot m)$  and  $O(c \cdot m^2)$ . The PSI in [38] has a parameter  $t$  that determines how many parties can collude with each other and must be set before the protocol’s execution, where  $t \in [2, m)$ . Its computation and communication complexities are  $O(c \cdot \kappa(m + t^2 - t(m + 1)))$  and  $O(c \cdot m \cdot \kappa)$  respectively.

Dong *et al.* introduced a “fair” two-party PSI [16], ensuring that both parties receive the result or neither does, even if a malicious party aborts prematurely during the protocol’s execution. This protocol utilises homomorphic encryption, zero-knowledge proofs, and polynomial set representation. The protocol’s computation and communication complexities are  $O(c^2)$  and  $O(c)$  respectively. Since then, various fair two-party PSIs have been proposed, e.g., in [13,15,14]. As of now, the fair PSI presented in citeDebnathD16 exhibits superior complexity and performance when compared to previous fair PSIs. It primarily relies on ElGamal encryption, verifiable encryption, and zero-knowledge proofs. The protocol’s computation and communication cost is  $O(c)$ . However, its overall overhead remains high, primarily due to its reliance on asymmetric key primitives, such as zero-knowledge proofs. To date, there has been no fair multi-party PSI in the literature. Justitia, our protocol, stands as the inaugural fair multi-party PSI.

Delegated PSIs leverage cloud computing for computation and/or storage, ensuring the confidentiality of the computation inputs and outputs from the cloud. These protocols can be further categorised into those that facilitate *one-off* and *repeated* delegation of PSI computation. The former, such as [28,30,48], cannot reuse their outsourced encrypted data and require clients to re-encode their data for each computation. The most efficient protocol among them is [28], designed for the two-party setting, with a computation and communication complexity of  $O(c)$ . In contrast, protocols supporting repeated PSI delegation let clients outsource their encrypted data to the cloud just once and subsequently perform an unlimited number of computations on the outsourced data.

The protocol in [1] is the first PSI that efficiently supports repeated delegation in the semi-honest model. It uses the polynomial representation of sets, pseudorandom function, and hash table. Its communication and computation complexities are  $O(h \cdot d^2)$  and  $O(h \cdot d)$  respectively, where  $h$  is the total number of bins in the hash table,  $d$  is a bin's capacity (often  $d = 100$ ), and  $h \cdot d$  is linear with  $c$ . Recently, a multi-party PSI that supports repeated delegation and efficient *updates* has been proposed in [2]. It is also in the semi-honest model. It imposes  $O(h \cdot d^2 \cdot m)$  and  $O(h \cdot d \cdot m)$  computation and communication costs respectively, during the PSI computation. It remains to be seen how a fair delegated PSI can be designed.

### 3 Preliminaries

#### 3.1 Notation Table

Table 1 summarises the main notations used in the paper.

Table 1: Notation Table.

Setting	Symbol	Description	Setting	Symbol	Description
Generic	$p$	Large prime number	Counter Collusion Contracts	$\mathcal{SC}_{rc}$	Prisoner's Contract
	$\mathbb{F}_p$	A finite field of prime order $p$		$\mathcal{SC}_{cc}$	Colluder's Contract
	$CL$	Set of all clients, $\{A_1, \dots, A_m, D\}$		$\mathcal{SC}_{tc}$	Traitor's Contract
	$D$	Dealer client		$\tilde{c}$	Server's cost for computing a task
	$A_m$	Buyer client		$\tilde{c}_h$	Auditor's cost for resolving disputes
	$m$	Total number of clients (excluding $D$ )		$\tilde{d}$	Deposit a server pays to get the job
	$H$	Hash function		$\tilde{w}$	Amount a server receives for completing the task
	$ S_{\cap} $	Intersection size		$(pk, sk)$	$\mathcal{SC}_{JUS}$ 's auditor's public-private key pair
	$S_{min}$	Smallest set's size	Justitia (JUS)	$\mathcal{SC}_{JUS}$	JUS's smart contract
	$S_{max}$	Largest set's size		$\omega, \omega', \rho$	Random poly. of degree $d$
	$ $	Divisible		$\gamma, \delta$	Random poly. of degree $d + 1$
	$\setminus$	Set subtraction		$\nu^{(C)}$	Blinded poly. sent by each $C$ to $\mathcal{SC}_{JUS}$
	$c$	Set's cardinality		$\phi$	Blinded poly. encoding the intersection
	$h$	Total number of bins in a hash table		$\chi$	Poly. sent to $\mathcal{SC}_{JUS}$ to identify misbehaving parties
	$d$	A bin's capacity		$L$	List of identified misbehaving parties
	$\lambda$	Security parameter		$\tilde{y}$	A portion of a party's deposit into $\mathcal{SC}_{JUS}$ transferred to honest clients if it misbehaves
	OLE	Oblivious Linear Evaluation		$mk$	Master key of PRF
	OLE <sup>+</sup>	Advanced OLE		$Q^{Init}$	Initiation predicate
	Com	Commitment algorithm of commitment		$Q^{Del}$	Delivery predicate
	Ver	Verification algorithm of commitment		$Q^{UF-A}$	UnFair-Abort predicate
	MT.genTree	Tree construction algorithm of Merkle tree		$Q^{F-A}$	Fair-Abort predicate
	MT.prove	Proof generation algorithm of Merkle tree	Anesidora (ANE)	$\mathcal{SC}_{ANE}$	ANE's smart contract
	MT.verify	Verification algorithm of Merkle tree		$\tilde{d}'$	Extractor's deposit
	CT	Coin tossing protocol		$\tilde{y}'$	Each client's deposit into $\mathcal{SC}_{JUS}$
	VOPR	Verifiable Oblivious Poly. Randomization		$\tilde{l}$	Reward a client earns for an intersection element
	ZSPA	Zero-sum Pseudorandom Values Agreement		$\tilde{r}$	Extractor's cost for extracting an intersection element
	ZSPA-A	ZSPA with an External Auditor		$\tilde{f}$	Shorthand for $\tilde{l}(m - 1)$
	$\mathcal{PST}^{FC}$	Multi-party PSI with Fair Compensation		$\tilde{v}$	Price a buyer pays for an intersection element $\tilde{v} = m \cdot \tilde{l} + 2\tilde{r}$
	$\mathcal{PST}^{FCR}$	Multi-party PSI with Fair Compensation and Reward		$mk'$	Another master key of PRF
	JUS	Protocol that realises $\mathcal{PST}^{FC}$		$ct_{mk}$	Encryption of $mk$ under $pk$
	ANE	Protocol that realises $\mathcal{PST}^{FCR}$		$Q^{Del}$	Delivery-with-Reward predicate
	PRF	Pseudorandom function		$Q^{UF-A}$	UnFair-Abort-with-Reward predicate
	PRP	Pseudorandom permutation			
	$gcd$	Greatest common divisor			
	$\epsilon$	Negligible function			

#### 3.2 Security Model

The paper employs a simulation-based secure computation model [24] to define and prove its protocols, addressing both active (malicious) and passive (semi-honest) adversaries. Definitions for both adversary types are outlined, assuming a static adversary and a secure (off-chain) communication channel between clients.

**Two-party Computation.** A two-party protocol  $\Gamma$  problem is captured by specifying a random process that maps pairs of inputs to pairs of outputs, one for each party. Such process is referred to as a functionality  $f : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^* \times \{0, 1\}^*$ , where  $f := (f_1, f_2)$ . For every input pair  $(x, y)$ , the output pair is a random variable  $(f_1(x, y), f_2(x, y))$ , such that the party with input  $x$  obtains  $f_1(x, y)$  while the party with input  $y$  receives  $f_2(x, y)$ . When  $f$  is deterministic, then  $f_1 = f_2$ . If  $f$  is asymmetric and only one party (say the first one) receives the result,  $f$  is defined as  $f := (f_1(x, y), \perp)$ .

**Security in the Presence of Passive Adversaries.** In this setting, a protocol is secure if whatever can be computed by a party in the protocol can be computed using its input and output only.

**Definition 1.** Let  $f$  be the deterministic functionality which was defined above. Protocol  $\Gamma$  security computes  $f$  in the presence of a static passive adversary if there exist polynomial time algorithms  $(\text{Sim}_1, \text{Sim}_2)$  such that:  $\{\text{Sim}_1(x, f_1(x, y))\}_{x, y} \stackrel{c}{=} \{\text{View}_1^\Gamma(x, y)\}_{x, y}$  and  $\{\text{Sim}_2(y, f_2(x, y))\}_{x, y} \stackrel{c}{=} \{\text{View}_2^\Gamma(x, y)\}_{x, y}$  where party  $i$ 's view (during the execution of  $\Gamma$ ) on input pair  $(x, y)$  is denoted by  $\text{View}_i^\Gamma(x, y)$  and equals  $(w, r^i, m_1^i, \dots, m_t^i)$ , where  $w \in \{x, y\}$  is the input of  $i^{\text{th}}$  party,  $r_i$  is the outcome of this party's internal random coin tosses, and  $m_j^i$  represents the  $j^{\text{th}}$  message this party receives.

**Security in the Presence of Active Adversaries.** In this setting, correctness is required beyond the possibility that a corrupted party may learn more than it should. To capture the threats, a protocol's security is analysed by comparing what an adversary can do in the real protocol to what it can do in an ideal scenario. This is formalised by considering an ideal computation involving an incorruptible Trusted Third Party (TTP) to whom the parties send their inputs and receive the output of the ideal functionality.

**Definition 2.** Let  $f$  be the two-party functionality defined above and  $\Gamma$  be a two-party protocol that computes  $f$ . Protocol  $\Gamma$  securely computes  $f$  with abort in the presence of static active adversaries if for every Probabilistic Polynomial Time (PPT) adversary  $\mathcal{A}$  for the real model, there exists a non-uniform PPT adversary (or simulator)  $\text{Sim}$  for the ideal model, such that for every  $i \in \{0, 1\}$ , it holds that:  $\{\text{Ideal}_{\text{Sim}(z), i}^f(x, y)\}_{x, y, z} \stackrel{c}{=} \{\text{Real}_{\mathcal{A}(z), i}^\Gamma(x, y)\}_{x, y, z}$ , where the ideal execution of  $f$  on inputs  $(x, y)$  and  $z$  is denoted by  $\text{Ideal}_{\mathcal{A}(z), i}^f(x, y)$  and is defined as the output pair of the honest party and  $\mathcal{A}$  from the ideal execution.

The real execution of  $\Gamma$  is denoted by  $\text{Real}_{\mathcal{A}(z), i}^\Gamma(x, y)$ , it is defined as the joint output of the parties engaging in the real execution of  $\Gamma$ , in the presence of  $\mathcal{A}$ .

### 3.3 Smart Contracts

Cryptocurrencies, such as Bitcoin [37] and Ethereum [46], go beyond merely providing a decentralised currency; they also facilitate computations on transactions. These cryptocurrencies allow for a specific computational logic to be encoded in a computer program known as a “smart contract”. As of now, Ethereum

stands as the foremost cryptocurrency framework that empowers users to define arbitrary smart contracts. Within this framework, contract code resides on the blockchain and is executed by all parties involved in maintaining the cryptocurrency. The correctness of program execution is ensured by the security of the underlying blockchain components. In the context of this work, **standard** public (Ethereum) smart contracts align with the requirements of our protocols.

### 3.4 Counter Collusion Smart Contracts

To enable a party, such as a client, to efficiently delegate computation to multiple potentially colluding third parties, like servers, Dong *et al.* [17] introduced two primary smart contracts: the “Prisoner’s Contract” ( $\mathcal{SC}_{\text{PC}}$ ) and the “Traitor’s Contract” ( $\mathcal{SC}_{\text{TC}}$ ).  $\mathcal{SC}_{\text{PC}}$  jointly signed by both the client and the servers and is designed to incentivise accurate computation. This contract mandates that each server must submit a deposit before the computation is delegated. Furthermore, it is equipped with an external auditor that can be invoked to identify any misbehaving server when they provide dissimilar results.

If a server behaves honestly, it is eligible to withdraw its deposit. However, if the auditor detects a cheating server, a portion of its deposit is transferred to the client. In the scenario where one server is honest while the other one cheats, the honest server receives a reward sourced from the deposit of the cheating server.

However, the dilemma, created by  $\mathcal{SC}_{\text{PC}}$  between the two servers, can be resolved if they are able to establish an enforceable promise, such as through a “Colluder’s Contract” ( $\mathcal{SC}_{\text{CC}}$ ). In this contract, one party, referred to as the “ringleader”, commits to paying a bribe to its counterpart if both parties engage in collusion and deliver an incorrect result to  $\mathcal{SC}_{\text{PC}}$ . To counter  $\mathcal{SC}_{\text{CC}}$ , Dong *et al.* proposed  $\mathcal{SC}_{\text{TC}}$ , which provides incentives for a colluding server to expose the other server and report the collusion without facing penalties from  $\mathcal{SC}_{\text{PC}}$ . In this study, we have made slight adjustments and employed these contracts. The relevant parameters for these contracts are outlined in 1. For a comprehensive description of the parameters and contracts, we direct readers to Appendix B.

### 3.5 Pseudorandom Function and Permutation

A pseudorandom function is a deterministic function that takes a key of length  $\lambda$  and an input; and outputs a value indistinguishable from that of a truly random function. In this paper, we use pseudorandom functions:  $\text{PRF} : \{0, 1\}^\lambda \times \{0, 1\}^* \rightarrow \mathbb{F}_p$ , where  $\log_2(p) = \lambda$  is the security parameter. In practice, a pseudorandom function can be obtained from an efficient block cipher [29]. The definition of a pseudorandom permutation,  $\text{PRP} : \{0, 1\}^\lambda \times \{0, 1\}^* \rightarrow \mathbb{F}_p$ , is similar to that of a pseudorandom function, with a difference, it is required  $\text{PRP}(k, \cdot)$  to be indistinguishable from a uniform permutation, instead of a uniform function.

### 3.6 Commitment Scheme

A commitment scheme involves a *sender* and a *receiver*. It includes two phases, *commit* and *open*. In the *commit* phase, the sender commits to a message  $x$  as  $\text{Com}(x, r) = \text{Com}_x$ , that involves a secret value,  $r$ . In the *open* phase, the sender sends the opening  $\tilde{x} := (x, r)$  to the receiver which verifies its correctness:

$\text{Ver}(\text{Com}_x, \tilde{x}) \stackrel{?}{=} 1$  and accepts if the output is 1. A commitment scheme must satisfy: (a) *hiding*: it is infeasible for an adversary to learn any information about the committed message  $x$ , until the commitment  $\text{com}$  is opened, and (b) *binding*: it is infeasible for a malicious sender to open a commitment  $\text{com}$  to different values than that was used in the commit phase. We refer readers to Appendix C for further details.

### 3.7 Hash Tables

A hash table is an array of bins, each capable of containing a set of elements, and is paired with a hash function. To insert an element, we initially compute the element’s hash and subsequently place the element into the bin corresponding to the computed hash value. In this paper, we ensure that the number of elements in each bin does not surpass a predefined capacity. By considering the maximum number of elements as  $c$  and the maximum size of a bin as  $d$ , we can calculate the number of bins, denoted as  $h$ , through an analysis of hash tables using the “balls into the bins” model [9]. In Appendix D, we provide an explanation of how the hash table parameters are configured.

### 3.8 Merkle Tree

A Merkle tree is a data structure that facilitates a concise commitment to a set of values or blocks, involving two parties: a prover and a verifier. The Merkle tree scheme includes three algorithms; namely, **MT.genTree**, **MT.prove**, and **MT.verify**. Briefly, the first algorithm constructs a Merkle tree on a set of blocks, the second generates a proof of a block’s (or set of blocks’) membership, and the third verifies the proof. Appendix E provides more details.

### 3.9 Polynomial Representation of Sets

The idea of using a polynomial to represent a set’s elements was proposed by Freedman *et al.* in [19]. Since then, the idea has been widely used, e.g., in [23,33]. In this representation, set elements  $S = \{s_1, \dots, s_d\}$  are defined over a finite field

$\mathbb{F}_p$  and set  $S$  is represented as a polynomial of form:  $\mathbf{p}(x) = \prod_{i=1}^d (x - s_i)$ , where

$\mathbf{p}(x) \in \mathbb{F}_p[X]$  and  $\mathbb{F}_p[X]$  is a polynomial ring. Often a polynomial,  $\mathbf{p}(x)$ , of degree  $d$  is represented in the “coefficient form” as follows:  $\mathbf{p}(x) = a_0 + a_1 \cdot x + \dots + a_d \cdot x^d$ .

The form  $\prod_{i=1}^d (x - s_i)$  is a special case of the coefficient form. As shown in [11,33],

for two sets  $S^{(A)}$  and  $S^{(B)}$  represented by polynomials  $\mathbf{p}_A$  and  $\mathbf{p}_B$  respectively, their product, which is polynomial  $\mathbf{p}_A \cdot \mathbf{p}_B$ , represents the set union, while their greatest common divisor,  $\text{gcd}(\mathbf{p}_A, \mathbf{p}_B)$ , represents the set intersection. For two polynomials  $\mathbf{p}_A$  and  $\mathbf{p}_B$  of degree  $d$ , and two random polynomials  $\gamma_A$  and  $\gamma_B$  of degree  $d$ , it is proven in [11,33] that:  $\theta = \gamma_A \cdot \mathbf{p}_A + \gamma_B \cdot \mathbf{p}_B = \mu \cdot \text{gcd}(\mathbf{p}_A, \mathbf{p}_B)$ , where  $\mu$  is a uniformly random polynomial, and polynomial  $\theta$  contains only information about the elements in  $S^{(A)} \cap S^{(B)}$ , and contains no information about other elements in  $S^{(A)}$  or  $S^{(B)}$ .

Given a polynomial  $\theta$  that encodes sets intersection, one can find the set elements in the intersection via one of the following approaches. First, via polynomial evaluation: the party who already has one of the original input sets, say

$\mathbf{p}_A$ , evaluates  $\theta$  at every element  $s_i$  of  $\mathbf{p}_A$  and considers  $s_i$  in the intersection if  $\mathbf{p}_A(s_i) = 0$ . Second, via polynomial root extraction: the party who does not have one of the original input sets, extracts the roots of  $\theta$ , which contain the roots of (i) random polynomial  $\mu$  and (ii) the polynomial that represents the intersection, i.e.,  $\gcd(\mathbf{p}_A, \mathbf{p}_B)$ . In this approach, to distinguish errors (i.e., roots of  $\mu$ ) from the intersection, PSIs in [1,33] use the “hash-based padding technique”. In this technique, every element  $u_i$  in the set universe  $\mathcal{U}$ , becomes  $s_i = u_i || \mathbf{H}(u_i)$ , where  $\mathbf{H}$  is a cryptographic hash function with a sufficiently large output size. Given a field’s arbitrary element,  $s \in \mathbb{F}_p$  and  $\mathbf{H}$ ’s output size  $|\mathbf{H}(\cdot)|$ , we can parse  $s$  into  $x_1$  and  $x_2$ , such that  $s = x_1 || x_2$  and  $|x_2| = |\mathbf{H}(\cdot)|$ . In a PSI that uses polynomial representation and this padding technique, after we extract each root of  $\theta$ , say  $s$ , we parse it into  $(x_1, x_2)$  and check  $x_2 \stackrel{?}{=} \mathbf{H}(x_1)$ . If the equation holds, then we consider  $s$  as an element of the intersection.

### 3.10 Horner’s Method

Horner’s method [18] allows for efficiently evaluating polynomials at a given point, e.g.,  $x_0$ . Specifically, given a polynomial of the form:  $\tau(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_d \cdot x^d$  and a point:  $x_0$ , one can efficiently evaluate the polynomial at  $x_0$  iteratively, in the following fashion:  $\tau(x_0) = a_0 + x_0(a_1 + x_0(a_2 + \dots + x_0(a_{d-1} + x_0 \cdot a_d) \dots))$ . Evaluating a polynomial of degree  $d$  naively requires  $d$  additions and  $\frac{(d^2+d)}{2}$  multiplications. However, using Horner’s method the evaluation requires only  $d$  additions and  $2d$  multiplications. We use this method in this paper.

### 3.11 Oblivious Linear Function Evaluation

Oblivious Linear function Evaluation (OLE) is a two-party protocol that involves a sender and receiver. In OLE, the sender has two inputs  $a, b \in \mathbb{F}_p$  and the receiver has a single input,  $c \in \mathbb{F}_p$ . The protocol allows the receiver to learn only  $s = a \cdot c + b \in \mathbb{F}_p$ , while the sender learns nothing. Ghosh *et al.* [21] proposed an efficient OLE that has  $O(1)$  overhead and involves mainly symmetric key operations. Later, in [22] an enhanced OLE, called OLE<sup>+</sup> was proposed. The latter ensures that the receiver cannot learn anything about the sender’s inputs, even if it sets its input to 0. In this paper, we use OLE<sup>+</sup>. We refer readers to Appendix F, for its construction.

### 3.12 Coin-Tossing Protocol

A Coin-Tossing protocol, CT, allows two mutually distrustful parties, say  $A$  and  $B$ , to jointly generate a single random bit. Formally, CT computes the functionality  $f_{\text{CT}}(in_A, in_B) \rightarrow (out_A, out_B)$ , which takes  $in_A$  and  $in_B$  as inputs of  $A$  and  $B$  respectively and outputs  $out_A$  to  $A$  and  $out_B$  to  $B$ , where  $out_A = out_B$ . A basic security requirement of a CT is that the resulting bit is (computationally) indistinguishable from a truly random bit. Two-party coin-tossing protocols can be generalised to *multi-party* coin-tossing ones to generate a *random string* (rather than a single bit). The overheads of multi-party coin-tossing protocols are often linear with the number of participants. In this paper, any secure multi-party CT that generates a random string can be used. For the sake of simplicity, we let a multi-party  $f_{\text{CT}}$  take  $m$  inputs and output a single value, i.e.,  $f_{\text{CT}}(in_1, \dots, in_m) \rightarrow out$ . We refer readers to Appendix G for further details.



## 4 Definition of Multi-party PSI with Fair Compensation

Now, we present the concept of multi-party PSI with Fair Compensation ( $\mathcal{PSI}^{\mathcal{FC}}$ ) which ensures that either all clients receive the result or that honest parties are financially compensated in the event of an unfair protocol abortion, wherein only dishonest parties gain access to the result.

In a  $\mathcal{PSI}^{\mathcal{FC}}$ , three types of parties are involved; namely, (1) a set of clients  $\{A_1, \dots, A_m\}$  potentially *malicious* (i.e., active adversaries), where all but one may collude with each other, (2) a non-colluding dealer,  $D$ , potentially semi-honest (i.e., a passive adversary), and (3) an auditor,  $Aud$ , potentially semi-honest, where all parties except  $Aud$  have input set. For simplicity, we assume that one can determine whether a given address belongs to  $Aud$ . The fundamental functionality computed by any multi-party PSI can be defined as  $f^{\text{PSI}}(S_1, \dots, S_{m+1}) \rightarrow \underbrace{(S_\cap, \dots, S_\cap)}_{m+1}$ , where  $S_\cap = S_1 \cap S_2, \dots, \cap S_{m+1}$ .

At a high level, the standard simulation-based paradigm ensures only privacy and output correctness [24]. To enhance the standard simulation-based model and accommodate additional security requirements (such as fairness, compensation, or transactions' correctness) researchers often follow these steps: (i) define a set of predicates and (ii) define a wrapper that encapsulates and parameterises the original paradigm with the defined predicates. This approach can be observed in [32,6,5]. Consequently, to formally define a  $\mathcal{PSI}^{\mathcal{FC}}$ , we parameterise  $f^{\text{PSI}}$  with four predicates,  $Q := (Q^{\text{Init}}, Q^{\text{Del}}, Q^{\text{UF-A}}, Q^{\text{F-A}})$ , which ensure that certain financial conditions are met. We borrow three of these predicates (i.e.,  $Q^{\text{Init}}, Q^{\text{Del}}, Q^{\text{UF-A}}$ ) from [32]. Nevertheless, we will (i) introduce an additional predicate  $Q^{\text{F-A}}$  and (ii) provide more formal accurate definitions of these predicates. Below, we provide a brief description of these predicates.

- $Q^{\text{Init}}$  defines the conditions under which a protocol realising  $\mathcal{PSI}^{\mathcal{FC}}$  should begin execution, i.e., when all set owners have sufficient deposit.
- $Q^{\text{Del}}$  determines the circumstances in which parties receive their output, specifically when honest parties receive their deposit back.
- $Q^{\text{UF-A}}$  specifies the conditions under which the simulator can force parties to abort if the adversary learns the output, i.e., when an honest party receives its deposit back along with a predefined amount of compensation.
- $Q^{\text{F-A}}$  specifies the conditions under which the simulator can compel parties to abort if the adversary receives no output, i.e., when honest parties receive their deposits back.

By requiring any protocol that realises  $\mathcal{PSI}^{\mathcal{FC}}$  to implement a wrapped version of  $f^{\text{PSI}}$  that includes  $Q$ , we will ensure that an honest set owner only aborts in an unfair manner if  $Q^{\text{UF-A}}$  returns 1, it only aborts in a fair manner if  $Q^{\text{F-A}}$  returns 1, and outputs a valid value if  $Q^{\text{Del}}$  returns 1. We refer readers to Appendix H for formal definitions of these predicates. Next, we present a formal definition of  $\mathcal{PSI}^{\mathcal{FC}}$ .

**Definition 3** ( $\mathcal{PSI}^{\mathcal{FC}}$ ). *Let  $f^{\text{PSI}}$  be the multi-party PSI functionality defined above. Then, protocol  $\Gamma$  realises  $f^{\text{PSI}}$  with  $Q$ -fairness in the presence of  $m - 1$*

static active-adversary clients (i.e.,  $A_js$ ) or a static passive dealer  $D$  or passive auditor  $Aud$ , if for every PPT adversary  $\mathcal{A}$  for the real model, there exists a PPT adversary (or simulator)  $\text{Sim}$  for the ideal model, such that for every  $I \in \{A_1, \dots, A_m, D, Aud\}$ , it holds that:  $\{\text{Ideal}_{\text{Sim}(z), I}^{\mathcal{W}(f^{PSI}, Q)}(S_1, \dots, S_{m+1})\}_{S_1, \dots, S_{m+1}, z} \stackrel{c}{=} \{\text{Real}_{\mathcal{A}(z), I}^r(S_1, \dots, S_{m+1})\}_{S_1, \dots, S_{m+1}, z}$ , where  $z$  is an auxiliary input given to  $\mathcal{A}$  and  $\mathcal{W}(f^{PSI}, Q)$  is a functionality that wraps  $f^{PSI}$  with predicates  $Q := (Q^{Init}, Q^{Del}, Q^{UF-A}, Q^{F-A})$ .

## 5 Other Subroutines Used in Justitia

In this section, we present three subroutines and a primitive that we have developed and use in the instantiation of  $\mathcal{PSI}^{\mathcal{F}^c}$ , known as Justitia.

### 5.1 Verifiable Oblivious Polynomial Randomisation (VOPR)

In the VOPR, two type of parties are involved, (i) a sender, potentially a passive adversary, and (ii) a receiver, potentially an active adversary. The protocol enables the receiver, with input polynomial  $\beta$  (of degree  $e'$ ), and the sender, with input random polynomials  $\psi$  (of degree  $e$ ) and  $\alpha$  (of degree  $e + e'$ ), to compute:  $\theta = \psi \cdot \beta + \alpha$ , satisfying the following conditions: (a) the receiver learns only  $\theta$  and gains no knowledge of the sender's input even when setting  $\beta = 0$ , (b) the sender gains no knowledge, and (c) protocol detects the receiver's misbehavior. Thus, the functionality that VOPR computes is defined as  $f^{\text{VOPR}}((\psi, \alpha), \beta) \rightarrow (\perp, \psi \cdot \beta + \alpha)$ .

We will use VOPR in Justitia for two main purposes: (a) to enable a party to re-randomise its counterpart's polynomial (representing its set) and (b) to impose a MAC-like structure on the randomised polynomial. This structure will allow a verifier to detect any modifications to VOPR's output.

Now, we will outline how we design VOPR without using any zero-knowledge proofs.<sup>2</sup> In the setup phase, both parties represent their input polynomials in the regular coefficient form; therefore, the sender's polynomials are defined as

$$\psi = \sum_{i=0}^e g_i \cdot x^i \text{ and } \alpha = \sum_{j=0}^{e+e'} a_j \cdot x^j \text{ and the receiver's polynomial is defined as } \beta = \sum_{i=0}^{e'} b_i \cdot x^i, \text{ where } b_i \neq 0. \text{ However, the sender computes each coefficient } a_j$$

(of polynomial  $\alpha$ ) as follows,  $a_j = \sum_{t=0}^{k=e'} a_{t,k}$ , where  $t + k = j$  and each  $a_{t,k}$  is a random value. For instance, if  $e = 4$  and  $e' = 3$ , then  $a_3 = a_{0,3} + a_{3,0} + a_{1,2} + a_{2,1}$ . Shortly, we explain why polynomial  $\alpha$  is constructed this way.

In the computation phase, to compute polynomial  $\theta$ , the two parties interactively multiply and add the related coefficients in a secure way using  $\text{OLE}^+$ . Specifically, for every  $j$  (where  $0 \leq j \leq e'$ ) the sender sends  $g_i$  and  $a_{i,j}$  to an

<sup>2</sup> Previously, Ghosh *et al.* [22] designed a protocol called Oblivious Polynomial Addition (OPA) to meet similar security requirements that we laid out above. But, as shown in [4], OPA is susceptible to several serious attacks.

instance of  $\text{OLE}^+$ , while the receiver sends  $b_j$  to the same instance, which returns  $c_{i,j} = g_i \cdot b_j + a_{i,j}$  to the receiver. This process is repeated for every  $i$ , where  $0 \leq i \leq e$ . Then, the receiver uses  $c_{i,j}$  values to construct the resulting polynomial,  $\theta = \psi \cdot \beta + \alpha$ . The reason the sender imposes the above structure on (the coefficients of)  $\alpha$  during the setup, is to enable the parties to securely compute  $\theta$  via  $\text{OLE}^+$ . Specifically, by imposing this structure (1) the sender can blind each product  $g_i \cdot b_j$  with random value  $a_{i,j}$  which is a component of  $\alpha$ 's coefficient and (2) the receiver can construct a result polynomial in the form  $\theta = \psi \cdot \beta + \alpha$ .

To check the correctness of the result, the sender selects and transmits a random value  $z$  to the receiver. The receiver computes  $\theta(z)$  and  $\beta(z)$  and sends these two values to the sender. The sender computes  $\psi(z)$  and  $\alpha(z)$  and then checks if equation  $\theta(z) = \psi(z) \cdot \beta(z) + \alpha(z)$  holds. It accepts the result if the check passes. Figure 1 describes VOPR in detail.

- **Input.**
  - *Public Parameters:* upper bound on input polynomials' degree:  $e$  and  $e'$ .
  - *Sender Input:* random polynomials:  $\psi = \sum_{i=0}^e g_i \cdot x^i$  and  $\alpha = \sum_{j=0}^{e+e'} a_j \cdot x^j$ , where
$$g_i \xleftarrow{\$} \mathbb{F}_p.$$
Each  $a_j$  has the form:  $a_j = \sum_{\substack{k=0 \\ t+k=j}}^{e'} a_{t,k}$ , such that  $t+k=j$  and  $a_{t,k} \xleftarrow{\$} \mathbb{F}_p$ .
  - *Receiver Input:* polynomial  $\beta = \beta_1 \cdot \beta_2 = \sum_{i=0}^{e'} b_i \cdot x^i$ , where  $\beta_1$  is a random polynomial of degree 1,  $\beta_2$  is an arbitrary polynomial of degree  $e' - 1$ , and  $b_i \neq 0$ .
- **Output.** The receiver gets  $\theta = \psi \cdot \beta + \alpha$ .
- 1. **Computation:**
  - (a) Sender and receiver together for every  $j$ ,  $0 \leq j \leq e'$ , invoke  $e+1$  instances of  $\text{OLE}^+$ . In particular,  $\forall j, 0 \leq j \leq e'$ : sender sends  $g_i$  and  $a_{i,j}$  while the receiver sends  $b_j$  to  $\text{OLE}^+$  that returns:  $c_{i,j} = g_i \cdot b_j + a_{i,j}$  to the receiver ( $\forall i, 0 \leq i \leq e$ ).
  - (b) The receiver sums component-wise values  $c_{i,j}$  that results in polynomial:
$$\theta = \psi \cdot \beta + \alpha = \sum_{\substack{i=0 \\ j=0}}^{e+e'} c_{i,j} \cdot x^{i+j}$$
- 2. **Verification:**
  - (a) Sender: picks a random (non-zero) value  $z$  and sends it to the receiver.
  - (b) Receiver: sends  $\theta_z = \theta(z)$  and  $\beta_z = \beta(z)$  to the sender.
  - (c) Sender: computes  $\psi_z = \psi(z)$  and  $\alpha_z = \alpha(z)$  and checks if equation  $\theta_z = \psi_z \cdot \beta_z + \alpha_z$  holds. If the equation holds, it concludes that the computation was performed correctly. Otherwise, it aborts.

Fig. 1: Verifiable Oblivious Polynomial Randomisation (VOPR)

Note that VOPR requires the sender to insert non-zero coefficients, i.e.,  $b_i \neq 0$  for all  $i, 0 \leq i \leq e'$ . If the sender includes a zero-coefficient, then it will only obtain a random value (due to  $\text{OLE}^+$ ), making it fail to pass VOPR's verification

phase. However, it is important to mention that this requirement does not affect the correctness of Justitia, as we will explain in Section 6.1 and Appendix N.3.

**Theorem 1.** *Let  $f^{\text{VOPR}}$  be the functionality defined above. If the enhanced OLE (i.e.,  $\text{OLE}^+$ ) is secure against malicious (or active) adversaries, then the Verifiable Oblivious Polynomial Randomisation (VOPR), presented in Figure 1, securely computes  $f^{\text{VOPR}}$  in the presence of (i) semi-honest sender and honest receiver or (ii) malicious receiver and honest sender.*

We refer readers to Appendix I for the proof of Theorem 1.

## 5.2 Zero-sum Pseudorandom Values Agreement Protocol (ZSPA)

The ZSPA allows  $m$  parties (the majority of which maybe malicious) to efficiently agree on (a set of vectors, where each  $i$ -th vector has)  $m$  pseudorandom values such that their sum equals zero. At a high level, the parties first sign a smart contract, register their accounts/addresses in it, and then run a coin-tossing protocol CT to agree on a key:  $k$ . Next, one of the parties generates  $m - 1$  pseudorandom values  $z_{i,j}$  (where  $1 \leq j \leq m - 1$ ) using key  $k$  and PRF. It sets the last value as the additive inverse of the sum of the values generated, i.e.  $z_{i,m} = -\sum_{j=1}^{m-1} z_{i,j}$  (similar

to the standard XOR-based secret sharing [43]). Then, it constructs a Merkle tree on top of the pseudorandom values and stores only the tree's root  $g$  and the key's hash value  $q$  in the smart contract. Then, each party (using the key) locally checks if the values (on the contract) have been constructed correctly; if so, then it sends a (signed) "approved" message to the contract which only accepts messages from registered parties. Hence, the functionality that ZSPA computes is defined as  $f^{\text{ZSPA}}(\underbrace{\perp, \dots, \perp}_m) \rightarrow ((k, g, q), \dots, (k, g, q))$ , where  $g$  is the root of the

Merkle tree built on the pseudorandom values  $z_{i,j}$ ,  $q$  is the hash value of the key used to generate the pseudorandom values, and  $m \geq 2$ . Figure 2 presents ZSPA in detail.

ZSPA will be used in Justitia to let clients  $\{A_1, \dots, A_m\}$  provably agree on sets of pseudorandom values. Every client uses the elements of these sets (as coefficients) to generate a pseudorandom polynomial such that when all clients' pseudorandom polynomials are added it would result in 0, due to ZSPA's property. Each of these polynomials will be used by a client to blind the messages it sends to the smart contract, to protect the privacy of the plaintext message (from *Aud*, *D*, and the public). To compute the sum of the plaintext messages, one can easily sum the blinded messages, which removes the blinding polynomials.

**Theorem 2.** *Let  $f^{\text{ZSPA}}$  be the functionality defined above. If CT is secure against a malicious adversary and the correctness of PRF, H, and Merkle tree hold, then ZSPA, in Figure 2, securely computes  $f^{\text{ZSPA}}$  in the presence of  $m - 1$  malicious adversaries.*

We refer readers to Appendix J for the proof of Theorem 2.

- Parties. A set of clients  $\{A_1, \dots, A_m\}$ .
  - Input.  $m$ : the total number of participants,  $adr$ : an address of a deployed smart contract which maintains the clients' addresses and accepts messages only from them, and  $b$ : the total number of vectors. Let  $b' = b - 1$ .
  - Output.  $k$ : a secret key that generates  $b$  vectors  $[z_{0,1}, \dots, z_{0,m}], \dots, [z_{b',1}, \dots, z_{b',m}]$ ,  $q$ : hash of the key,  $g$ : a Merkle tree's root, and a vector of (signed) messages.
1. **Coin-tossing.**  $CT(in_1, \dots, in_m) \rightarrow k$ .  
All participants run a coin-tossing protocol to agree on PRF's key,  $k$ .
  2. **Encoding.**  $Encode(k, m) \rightarrow (g, q)$ .  
One of the parties takes the following steps:
    - (a) for every  $i$  ( $0 \leq i \leq b'$ ), generates  $m$  pseudorandom values:
 
$$\forall j, 1 \leq j \leq m - 1 : z_{i,j} = \text{PRF}(k, i || j), \quad z_{i,m} = - \sum_{j=1}^{m-1} z_{i,j}$$
    - (b) constructs a Merkle tree on top of all pseudorandom values,  $\text{MT.genTree}(z_{0,1}, \dots, z_{b',m}) \rightarrow g$ .
    - (c) sends the tree's root:  $g$ , and the key's hash:  $q = H(k)$  to  $adr$ .
  3. **Verification.**  $\text{Verify}(k, g, q, m, flag) \rightarrow (a, s)$ . Each  $A_t$  checks if, all  $z_{i,j}$  values, the root  $g$ , and key's hash  $q$  have been correctly generated, by retaking step 2.
    - If the checks pass, then it sets  $a = 1$  and
      - if  $flag = A$ , where  $A \in \{A_1, \dots, A_m\}$ , then it sets  $s$  to a (signed) “ap-proved” message, and sends  $s$  to  $adr$
      - if  $flag = A$ , where  $A \notin \{A_1, \dots, A_m\}$ , then it sets  $s = \perp$ .
    - If the checks do not pass, it aborts by returning  $a = 0$  and  $s = \perp$ .

Fig. 2: Zero-sum Pseudorandom Values Agreement (ZSPA).  $flag$  lets an external auditor locally run the verification without having to send a message to the smart contract.

### 5.3 ZSPA's Extension: ZSPA with an External Auditor (ZSPA-A)

Now, we present an extension of ZSPA, called ZSPA-A, which allows a (trusted) third-party auditor,  $Aud$ , to assist in identifying misbehaving clients within the ZSPA and generate a vector of random polynomials with a specific structure. Specifically, the functionality that ZSPA-A computes is defined as  $f^{\text{ZSPA-A}}(\underbrace{\perp, \dots, \perp}_m, b,$

$\zeta) \rightarrow ((\underbrace{(k, g, q), \dots, (k, g, q)}_m), \mu^{(1)}, \dots, \mu^{(m)})$ , where  $g$  is the root of the Markle tree

built on the pseudorandom values  $z_{i,j}$ ,  $q$  is the hash value of the key used to generate the pseudorandom values,  $m \geq 2$ ,  $b$  is constant value,  $\zeta$  is a random polynomial of degree 1, each  $\mu^{(j)}$  has the form  $\mu^{(j)} = \zeta \cdot \xi^{(j)} - \tau^{(j)}$ ,  $\xi^{(j)}$  is a random polynomial of degree  $b - 2$ , and  $\tau^{(j)} = \sum_{i=0}^{b-1} z_{i,j} \cdot x^i$ .

Informally, ZSPA-A ensures that misbehaving parties are always detected, except with a negligible probability.  $Aud$  of this protocol will be invoked by Justitia when the smart contract of Justitia detects that a combination of the messages sent by the clients is not well-formed. Later, in Justitia's proof, we will demonstrate that even a *semi-honest*  $Aud$ , who observes all messages send

by clients to Justitia's smart contracts, cannot learn anything about their set elements. We present ZSPA-A in Figure 3.

- **Parties.** A set of clients  $\{A_1, \dots, A_m\}$  and an external auditor,  $Aud$ .
  - **Input.**  $m$ : the total number of participants (excluding the auditor),  $\zeta$ : a random polynomial of degree 1,  $b$ : the total number of vectors, and  $adr$ : a deployed smart contract's address. Let  $b' = b - 1$ .
  - **Output of each  $A_j$ .**  $k$ : a secret key that generates  $b$  vectors  $[z_{0,1}, \dots, z_{0,m}], \dots, [z_{b',1}, \dots, z_{b',m}]$  of pseudorandom values,  $h$ : hash of the key,  $g$ : a Merkle tree's root, and a vector of (signed) messages.
  - **Output of  $Aud$ .**  $L$ : a list of misbehaving parties' indices, and  $\vec{\mu}$ : a vector of random polynomials.
1. **ZSPA invocation.**  $\text{ZSPA}(\perp, \dots, \perp) \rightarrow ((k, g, q), \dots, (k, g, q))$ .  
All parties in  $\{A_1, \dots, A_m\}$  call the same instance of ZSPA, which results in  $(k, g, q), \dots, (k, g, q)$ .
  2. **Auditor computation.**  $\text{Audit}(m, \vec{k}, q, \zeta, b, g) \rightarrow (L, \vec{\mu})$ .  
 $Aud$  takes the below steps. Note, each  $k_j \in k$  is given by  $A_j$ . An honest party's input,  $k_j$ , equals  $k$ , where  $1 \leq j \leq m$ .
    - (a) runs the checks in the verification phase (i.e., Phase 3) of ZSPA for every  $j$ , i.e.,  $\text{Verify}(k_j, g, q, m, auditor) \rightarrow (a_j, s)$ .
    - (b) appends  $j$  to  $L$ , if any checks fails, i.e., if  $a_j = 0$ . In this case, it skips the next two steps for the current  $j$ .
    - (c) For every  $i$  (where  $0 \leq i \leq b'$ ), it recomputes  $m$  pseudorandom values:  $\forall j, 1 \leq j \leq m - 1 : z_{i,j} = \text{PRF}(k, i || j), \quad z_{i,m} = - \sum_{j=1}^{m-1} z_{i,j}$ .
    - (d) generates polynomial  $\mu^{(j)}$  as:  $\mu^{(j)} = \zeta \cdot \xi^{(j)} - \tau^{(j)}$ , where  $\xi^{(j)}$  is a random polynomial of degree  $b' - 1$  and  $\tau^{(j)} = \sum_{i=0}^{b'} z_{i,j} \cdot x^i$ . By this step, a vector  $\vec{\mu}$  containing at most  $m$  polynomials is generated.
    - (e) returns list  $L$  and  $\vec{\mu}$ .

Fig. 3: ZSPA with an external auditor (ZSPA-A)

**Theorem 3.** *If ZSPA is secure, H is second-preimage resistant, and the correctness of PRF, H, and Merkle tree hold, then ZSPA-A securely computes  $f^{\text{ZSPA-A}}$  in the presence of  $m - 1$  malicious adversaries.*

We refer readers to Appendix K for the proof of Theorem 3.

#### 5.4 Unforgeable Polynomials

In this section, we introduce the new concept of unforgeable polynomials. Informally, an unforgeable polynomial possesses a secret factor. To verify that an unforgeable polynomial has not been tampered with, a verifier can check whether the polynomial is divisible by the secret factor. To transform an arbitrary polynomial  $\pi$  of degree  $d$  into an unforgeable polynomial  $\theta$ , we can follow these steps: (i) select two secret random polynomials  $(\zeta, \gamma)$  and (ii) calculate  $\theta = \zeta \cdot \pi + \gamma \bmod p$ , where  $\deg(\zeta) = 1$ , and  $\deg(\gamma) = d + 1$ .

To determine whether  $\theta$  has been tampered with, a verifier (given  $\theta, \gamma$ , and  $\zeta$ ) can check if  $\theta - \gamma$  is divisible by  $\zeta$ . Informally, the security of an unforgeable polynomial asserts that an adversary (who is unaware of the two secret random polynomials) cannot tamper with an unforgeable polynomial without being detected, except with a negligible probability. Below, we formally state it.

**Theorem 4 (Unforgeable Polynomial).** *Let polynomials  $\zeta$  and  $\gamma$  be three secret uniformly random polynomials (i.e.,  $\zeta, \gamma \xleftarrow{\$} \mathbb{F}_p[x]$ ), polynomial  $\pi$  be an arbitrary polynomial,  $\deg(\zeta) = 1$ ,  $\deg(\gamma) = d + 1$ ,  $\deg(\pi) = d$ , and  $p$  be a  $\lambda$ -bit prime number. Also, let polynomial  $\theta$  be defined as  $\theta = \zeta \cdot \pi + \gamma \bmod p$ . Given  $(\theta, \pi)$ , the probability that an adversary (which does not know  $\zeta$  and  $\gamma$ ) can forge  $\theta$  to an arbitrary polynomial  $\delta$  such that  $\delta \neq \theta$ ,  $\deg(\delta) = \text{const}(\lambda)$ , and  $\zeta$  divides  $\delta - \gamma$  is negligible in the security parameter  $\lambda$ , i.e.,  $\Pr[\zeta \mid (\delta - \gamma)] \leq \epsilon(\lambda)$ .*

An interesting feature of an unforgeable polynomial is that the verifier can perform the check without needing knowledge of the original polynomial  $\pi$ . Another appealing aspect of the unforgeable polynomial is its support for *linear combination* and, consequently, *batch verification*. Specifically, to transform  $n$  arbitrary polynomials  $[\pi_1, \dots, \pi_n]$  into unforgeable polynomials, one can construct  $\theta_i = \zeta \cdot \pi_i + \gamma_i \bmod p$ , where  $\forall i, 1 \leq i \leq n$ .

To verify the integrity of all polynomials  $[\theta_1, \dots, \theta_n]$ , a verifier can (i) compute their sum  $\chi = \sum_{i=1}^n \theta_i$  and (ii) check whether  $\chi - \sum_{i=1}^n \gamma_i$  is divisible by  $\zeta$ . Informally, the security of an unforgeable polynomial states that an adversary (who is unaware of the two secret random polynomials for each  $\theta_i$ ) cannot tamper with any subset of the unforgeable polynomials without being detected, except with a negligible probability. We formally state it, below.

**Theorem 5 (Unforgeable Polynomials' Linear Combination).** *Let polynomial  $\zeta$  be a secret polynomial picked uniformly at random; also, let  $\vec{\gamma} = [\gamma_1, \dots, \gamma_n]$  be a vector of secret uniformly random polynomials (i.e.,  $\zeta, \gamma_i \xleftarrow{\$} \mathbb{F}_p[x]$ ),  $\vec{\pi} = [\pi_1, \dots, \pi_n]$  be a vector of arbitrary polynomials,  $\deg(\zeta) = 1$ ,  $\deg(\gamma_i) = d + 1$ ,  $\deg(\pi_i) = d$ ,  $p$  be a  $\lambda$ -bit prime number, and  $1 \leq i \leq n$ . Moreover, let polynomial  $\theta_i$  be defined as  $\theta_i = \zeta \cdot \pi_i + \gamma_i \bmod p$ , and  $\vec{\theta} = [\theta_1, \dots, \theta_n]$ . Given  $(\vec{\theta}, \vec{\pi})$ , the probability that an adversary (which does not know  $\zeta$  and  $\vec{\gamma}$ ) can forge  $t$  polynomials, without loss of generality, say  $\theta_1, \dots, \theta_t \in \vec{\theta}$  to arbitrary polynomials  $\delta_1, \dots, \delta_t$  such that  $\sum_{j=1}^t \delta_j \neq \sum_{j=1}^t \theta_j$ ,  $\deg(\delta_j) = \text{const}(\lambda)$ , and  $\zeta$  divides  $(\sum_{j=1}^t \delta_j + \sum_{j=t+1}^n \theta_j - \sum_{j=1}^n \gamma_j)$  is negligible in the security parameter  $\lambda$ , i.e.,  $\Pr[\zeta \mid (\sum_{j=1}^t \delta_j + \sum_{j=t+1}^n \theta_j - \sum_{j=1}^n \gamma_j)] \leq \epsilon(\lambda)$ .*

Appendix L presents the proofs of Theorems 4 and 5. In Justitia, we will use unforgeable polynomials to enable a smart contract to efficiently verify the integrity of the polynomials sent by clients, ensuring that they are indeed the outputs of VOPR.

## 6 Concrete Construction of $\mathcal{PSI}^{\mathcal{FC}}$

### 6.1 An Overview of Justitia (JUS)

At a very high level, Justitia (JUS) works as follows. First, each client encodes its set elements into a polynomial. All clients sign a smart contract  $\mathcal{SC}_{\text{JUS}}$  and deposit a predefined amount of coins into it. One of the clients as a dealer,  $D$ , randomises the rest of the clients' polynomials and imposes a certain structure to their polynomials. The clients also randomise  $D$ 's polynomials. Next, all clients send their randomised polynomials to  $\mathcal{SC}_{\text{JUS}}$ .

$\mathcal{SC}_{\text{JUS}}$  combines all polynomials and checks whether the resulting polynomial still has the structure imposed by  $D$ . If it concludes that the resulting polynomial does not have the structure, then it invokes an auditor,  $Aud$ , to identify misbehaving clients and penalise them. Nevertheless, if the resulting polynomial has the structure, then  $\mathcal{SC}_{\text{JUS}}$  outputs an encoded polynomial and refunds the clients' deposits. In this case, all clients can use the resulting polynomial (output by  $\mathcal{SC}_{\text{JUS}}$ ) to locally find the intersection. Figure 6 in Appendix M outlines the interaction between parties.

One of the novelties of JUS is a lightweight verification mechanism which lets a smart contract  $\mathcal{SC}_{\text{JUS}}$  efficiently verify the correctness of the clients' messages without being able to learn/reveal the clients' set elements. To achieve this,  $D$  randomises each client's polynomials and constructs unforgeable polynomials on the randomised polynomials. If any client modifies an unforgeable polynomial that it receives and sends the modified polynomial to  $\mathcal{SC}_{\text{JUS}}$ , then  $\mathcal{SC}_{\text{JUS}}$  would detect it, by checking whether the sum of all clients' (unforgeable) polynomials is divisible by a certain polynomial of degree 1. The verification is lightweight because: (i) it does not use any public key cryptography (often computationally expensive), (ii) it needs to perform only polynomial division, and (iii) it can perform batch verification, i.e., it sums all clients randomised polynomials and then checks the result's correctness.

Now, we describe JUS in more detail. First, all clients sign and deploy  $\mathcal{SC}_{\text{JUS}}$ . Each of them put a certain amount of deposit into it. Then, they together run CT to agree on a key,  $mk$ , that will be used to generate a set of blinding polynomials to hide the final result from the public. Next, each client locally maps its set elements to a hash table and represents the content of each hash table's bin as a polynomial,  $\pi$ . After that, for each bin, the following steps are taken. All clients, except  $D$ , engage in ZSPA-A to agree on a set of pseudorandom blinding factors such that their sum is zero.

Then,  $D$  randomises and constructs an unforgeable polynomial on each client's polynomial,  $\pi$ . To do that,  $D$  and every other client, say  $C$ , engage in VOPR that returns to  $C$  a polynomial.  $D$  and every  $C$  invoke VOPR again to randomise  $D$ 's polynomial. VOPR returns to  $C$  another unforgeable polynomial. Note, the output of VOPR reveals nothing about any client's original polynomial  $\pi$ , as this polynomial has been blinded with another secret random polynomial by  $D$ , during the execution of VOPR. Each  $C$  sums the two polynomials blinds the result (using the output of ZSPA-A), and sends it to  $\mathcal{SC}_{\text{JUS}}$ . After all of the clients send their input polynomials to  $\mathcal{SC}_{\text{JUS}}$ ,  $D$  sends to  $\mathcal{SC}_{\text{JUS}}$  a *switching polynomial* that will



let  $\mathcal{SC}_{\text{JUS}}$  obviously switch the secret blinding polynomials used by  $D$  (during the execution of  $\text{VOPR}$ ) to blind each client's original polynomial  $\pi$  to another blinding polynomial that all clients can generate themselves, by using key  $mk$ .

Next,  $D$  sends to  $\mathcal{SC}_{\text{JUS}}$  a secret polynomial,  $\zeta$ , that will let  $\mathcal{SC}_{\text{JUS}}$  check unforgeable polynomials' correctness. Then,  $\mathcal{SC}_{\text{JUS}}$  sums all clients' polynomials and checks if  $\zeta$  can divide the sum.  $\mathcal{SC}_{\text{JUS}}$  accepts the clients' inputs if the polynomial divides the sum; otherwise, it invokes  $Aud$  to identify misbehaving parties. In this case, all honest parties' deposit is refunded to them and the deposit of misbehaving parties is distributed among the honest ones as well. If all clients behave honestly, then each client can locally find the intersection. To do that, it uses  $mk$  to remove the blinding polynomial from the sum (that the contract generated), evaluates the unblinded polynomial at each of its set elements and considers an element in the intersection if the evaluation is zero.

## 6.2 Detailed Description of JUS.

Now, we explain how JUS exactly works (see Table 1 in Appendix 3.1 for a description of the main notations used).

1. All clients in  $CL = \{A_1, \dots, A_m, D\}$  sign a smart contract:  $\mathcal{SC}_{\text{JUS}}$  and deploy it to a blockchain. All clients get the deployed contract's address. Also, all clients engage in  $\text{CT}$  to agree on a secret master key,  $mk$ .
2. Each client in  $CL$  builds a hash table,  $\text{HT}$ , and inserts the set elements into it, i.e.,  $\forall i : \text{H}(s_i) = \text{indx}$ , then  $s_i \rightarrow \text{HT}_{\text{indx}}$ . It pads every bin with random dummy elements to  $d$  elements (if needed). Then, for every bin, it constructs a polynomial whose roots are the bin's content:  $\pi = \prod_{i=1}^d (x - s'_i)$ , where  $s'_i$  is either  $s_i$  or a random value.
3. Every client  $C$  in  $CL \setminus D$ , for every bin, agree on  $b = 3d + 3$  vectors of pseudorandom blinding factors:  $z_{i,j}$ , such that the sum of each vector elements is zero, i.e.,  $\sum_{j=1}^m z_{i,j} = 0$ , where  $0 \leq i \leq b-1$ . To do that, they participate in step 1 of  $\text{ZSPA-A}$ . By the end of this step, for each bin, they agree on a secret key  $k$  (that will be used to generate the zero-sum values) as well as two values stored in  $\mathcal{SC}_{\text{JUS}}$ , i.e.,  $q$ : the key's hash value and  $g$ : a Merkle tree's root. After time  $t_1$ ,  $D$  ensures that all other clients have agreed on the vectors (i.e., all provided "approved" to the contract). If the check fails, it halts.
4. Each client in  $CL$  deposits  $\ddot{y} + \ddot{c}h$  amount to  $\mathcal{SC}_{\text{JUS}}$ . After time  $t_2$ ,  $\mathcal{SC}_{\text{JUS}}$  checks if in total  $(\ddot{y} + \ddot{c}h) \cdot (m + 1)$  amount has been deposited. If the check fails, it refunds the clients' deposit is refunded and halts.
5.  $D$  picks a random polynomial  $\zeta \xleftarrow{\$} \mathbb{F}_p[X]$  of degree 1, for each bin. It, for each client  $C$ , allocates to each bin two random polynomials:  $\omega^{(D,C)}, \rho^{(D,C)} \xleftarrow{\$} \mathbb{F}_p[X]$  of degree  $d$ , and two random polynomials:  $\gamma^{(D,C)}, \delta^{(D,C)} \xleftarrow{\$} \mathbb{F}_p[X]$  of degree  $3d+1$ . Also, each client  $C$ , for each bin, picks two random polynomials:  $\omega^{(C,D)}, \rho^{(C,D)} \xleftarrow{\$} \mathbb{F}_p[X]$  of degree  $d$ , and checks polynomials  $\omega^{(C,D)} \cdot \pi^{(C)}$  and  $\rho^{(C,D)}$  do not contain zero coefficients.
6.  $D$  randomises other clients' polynomials. To do so, for every bin, it invokes an instance of  $\text{VOPR}$  (presented in Fig. 1) with each client  $C$ ; where  $D$  sends

$\zeta \cdot \omega^{(D,C)}$  and  $\gamma^{(D,C)}$ , while client  $C$  sends  $\omega^{(C,D)} \cdot \pi^{(C)}$  to VOPR. Each client  $C$ , for every bin, receives a blind polynomial of the following form:

$$\theta_1^{(C)} = \zeta \cdot \omega^{(D,C)} \cdot \omega^{(C,D)} \cdot \pi^{(C)} + \gamma^{(D,C)}$$

from VOPR. If any party aborts, the deposit would be refunded to all parties.

7. Each client  $C$  randomises  $D$ 's polynomial. To do that, each client  $C$ , for each bin, invokes an instance of VOPR with  $D$ , where each client  $C$  sends  $\rho^{(C,D)}$ , while  $D$  sends  $\zeta \cdot \rho^{(D,C)} \cdot \pi^{(D)}$  and  $\delta^{(D,C)}$  to VOPR. Every client  $C$ , for each bin, receives a blind polynomial of the following form:

$$\theta_2^{(C)} = \zeta \cdot \rho^{(D,C)} \cdot \rho^{(C,D)} \cdot \pi^{(D)} + \delta^{(D,C)}$$

from VOPR. If a party aborts, all parties get their deposit back.

8. Each client  $C$ , for every bin, masks the sum of polynomials  $\theta_1^{(C)}$  and  $\theta_2^{(C)}$  using the blinding factors:  $z_{i,c}$ , generated in step 3. Specifically, it computes the following blind polynomial (for every bin):  $\nu^{(C)} = \theta_1^{(C)} + \theta_2^{(C)} + \tau^{(C)}$ , where  $\tau^{(C)} = \sum_{i=0}^{3d+2} z_{i,c} \cdot x^i$ . Next, it sends all  $\nu^{(C)}$  to  $\mathcal{SC}_{\text{JUS}}$ . If any party aborts, the deposit would be refunded to all parties.
9.  $D$  ensures all clients sent their inputs to  $\mathcal{SC}_{\text{JUS}}$ . If the check fails, it halts and the deposit is refunded to all parties. It allocates a pseudorandom polynomial  $\gamma'$  of degree  $3d$ , to each bin. To do so, it uses  $mk$  to derive a key for each bin:  $k_{\text{indx}} = \text{PRF}(mk, \text{indx})$  and uses  $k_{\text{indx}}$  to generate  $3d + 1$  pseudorandom coefficients  $g_{j, \text{indx}} = \text{PRF}(k_{\text{indx}}, j)$  where  $0 \leq j \leq 3d$ . Also, for each bin, it allocates a random polynomial  $\omega'^{(D)}$  of degree  $d$ .

10.  $D$ , for every bin, computes a polynomial of the form:

$$\nu^{(D)} = \zeta \cdot \omega'^{(D)} \cdot \pi^{(D)} - \sum_{C=A_1}^{A_m} (\gamma^{(D,C)} + \delta^{(D,C)}) + \zeta \cdot \gamma'$$

It sends to  $\mathcal{SC}_{\text{JUS}}$  polynomials  $\nu^{(D)}$  and  $\zeta$ , for each bin.

11.  $\mathcal{SC}_{\text{JUS}}$  takes the following steps:

- (a) for every bin, sums all related polynomials provided by all clients in  $\bar{P}$  as follows:  $\phi = \nu^{(D)} + \sum_{C=A_1}^{A_m} \nu^{(C)} = \zeta \cdot \left( \omega'^{(D)} \cdot \pi^{(D)} + \sum_{C=A_1}^{A_m} (\omega^{(D,C)} \cdot \omega^{(C,D)} \cdot \pi^{(C)}) + \pi^{(D)} \cdot \sum_{C=A_1}^{A_m} (\rho^{(D,C)} \cdot \rho^{(C,D)}) + \gamma' \right)$
- (b) checks whether, for every bin,  $\zeta$  divides  $\phi$ . If the check passes, it sets  $Flag = \text{True}$ . Otherwise, it sets  $Flag = \text{False}$ .
12. If  $Flag = \text{True}$ , then the following steps are taken:
  - (a)  $\mathcal{SC}_{\text{JUS}}$  sends back each party's deposit, i.e.,  $\ddot{y} + \ddot{c}h$  amount.
  - (b) each client (given  $\zeta$  and  $mk$ ) finds the elements in the intersection as follows.
    - i. derives a bin's pseudorandom polynomial,  $\gamma'$ , from  $mk$ .
    - ii. removes the blinding polynomial from each bin's polynomial:  $\phi' = \phi - \zeta \cdot \gamma'$
    - iii. evaluates each bin's unblinded polynomial at every element  $s_i$  belonging to that bin and considers the element in the intersection if the evaluation is zero: i.e.,  $\phi'(s_i) = 0$ .
13. If  $Flag = \text{False}$ , then the following steps are taken.

- (a) *Aud* asks every client  $C$  to send to it the PRF's key (generated in step 3), for every bin. It inserts the keys to  $\vec{k}$ . It generates a list  $\bar{L}$  initially empty. Then, for every bin, *Aud* takes step 2 of ZSPA-A, i.e., invokes  $\text{Audit}(m, \vec{k}, q, \zeta, 3d+3, g) \rightarrow (L, \vec{\mu})$ . Every time it invokes  $\text{Audit}$ , it appends the elements of returned  $L$  to  $\bar{L}$ . *Aud* for each bin sends  $\vec{\mu}$  to  $\mathcal{SC}_{\text{JUS}}$ . It also sends to  $\mathcal{SC}_{\text{JUS}}$  the list  $\bar{L}$  of all misbehaving clients detected so far.
- (b) to help identify further misbehaving clients,  $D$  takes the following steps, for each bin of client  $C$  whose ID is not in  $\bar{L}$ .
- i. computes polynomial  $\chi^{(D,C)}$  as follows.  

$$\chi^{(D,C)} = \zeta \cdot \eta^{(D,C)} - (\gamma^{(D,C)} + \delta^{(D,C)})$$
where  $\eta^{(D,C)}$  is a fresh random polynomial of degree  $3d+1$ .
  - ii. sends polynomial  $\chi^{(D,C)}$  to  $\mathcal{SC}_{\text{JUS}}$ .
- Note, if  $\bar{L}$  contains all clients' IDs, then  $D$  does not need to take the above steps 13(b)i and 13(b)ii.
- (c)  $\mathcal{SC}_{\text{JUS}}$ , takes the following steps to check if the client misbehaved, for each bin of client  $C$  whose ID is not in  $\bar{L}$ .
- i. computes polynomial  $\iota^{(C)}$  as follows:  

$$\iota^{(C)} = \chi^{(D,C)} + \nu^{(C)} + \mu^{(C)} = \zeta \cdot (\eta^{(D,C)} + \omega^{(D,C)} \cdot \omega^{(C,D)} \cdot \pi^{(C)} + \rho^{(D,C)} \cdot \rho^{(C,D)} \cdot \pi^{(D)} + \xi^{(C)})$$
where  $\mu^{(C)} \in \vec{\mu}$  was sent to  $\mathcal{SC}_{\text{JUS}}$  by *Aud* in step 13a.
  - ii. checks if  $\zeta$  divides  $\iota^{(C)}$ . If the check fails, it appends the client's ID to a list  $L'$ .
- If  $\bar{L}$  contains all clients' IDs, then  $\mathcal{SC}_{\text{JUS}}$  does not take the above two steps.
- (d)  $\mathcal{SC}_{\text{JUS}}$  refunds the honest parties' deposit. Also, it retrieves the total amount of  $\check{c}h$  from the deposit of dishonest clients (i.e., those clients whose IDs are in  $\bar{L}$  or  $L'$ ) and sends it to *Aud*. It also splits the remaining deposit of the misbehaving parties among the honest ones. Thus, each honest client receives  $\check{y} + \check{c}h + \frac{m' \cdot (\check{y} + \check{c}h) - \check{c}h}{m - m'}$  amount in total, where  $m'$  is the total number of misbehaving parties.

We refer readers to Appendix N for a more in-depth discussion of *Justitia* which encompasses (i) the inadequacy of strawman approaches such as using a server-aided PSI or charging the buyer a flat fee, (ii) instructions on how to eliminate the check in step 5, (iii) an outline of the primary challenges we addressed during *Justitia* design, and (iv) choice of concrete (security) parameters.

Next, we present a theorem that formally states the security of JUS.

**Theorem 6.** *Let polynomials  $\zeta$ ,  $\omega$ , and  $\gamma$  be three secret uniformly random polynomials. If  $\theta = \zeta \cdot \omega \cdot \pi + \gamma \bmod p$  is an unforgeable polynomial (w.r.t. Theorem 4), ZSPA-A, VOPR, PRF, and smart contracts are secure, then JUS securely realises  $f^{\text{PSI}}$  with  $Q$ -fairness (w.r.t. Definition 3) in the presence of  $m-1$  active-adversary clients (i.e.,  $A_j$ s) or a passive dealer client, passive auditor, or passive public.*

We refer readers to Appendix O for the proof of Theorem 6.

## 7 Definition of Multi-party PSI with Fair Compensation and Reward

In this section, we upgrade  $\mathcal{PSI}^{\mathcal{F}^C}$  to “multi-party PSI with Fair Compensation and Reward” ( $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$ ), which (in addition to offering the features of  $\mathcal{PSI}^{\mathcal{F}^C}$ ) lets honest clients who contribute their set receive a reward from a buyer who initiates the PSI computation.

In  $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$ , there are (1) a set of clients  $\{A_1, \dots, A_m\}$  a subset of which can be active adversaries and may collude with each other, (2) a non-colluding dealer client,  $D$ , potentially semi-honest, and (3) an auditor  $Aud$  potentially semi-honest, where all clients have input set. Also, there are two “extractor” clients, say  $A_1$  and  $A_2$ , where  $(A_1, A_2) \in \{A_1, \dots, A_m\}$ . These extractor clients volunteer to extract the (encoded) elements of the intersection and send them to a public bulletin board, i.e., a smart contract. In return, they will be paid. We assume these two extractors act rationally only when they want to carry out the paid task of extracting the intersection and reporting it to the smart contract, so they can maximise their profit.<sup>3</sup> For simplicity, we let client  $A_m$  be the buyer, i.e., the party which initiates the PSI computation and is interested in the result.

The formal definition of  $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$  is built upon the definition of  $\mathcal{PSI}^{\mathcal{F}^C}$ ; nevertheless, in  $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$ , we ensure that honest non-buyer clients receive a *reward* for participating in the protocol and revealing a portion of their inputs deduced from the result. We: (i) upgrade  $Q^{\text{Del}}$  to  $Q_R^{\text{Del}}$  to ensure that when honest clients receive the result, an honest non-buyer client receives its deposit back plus a reward and a buyer client receives its deposit back minus the paid reward, and (ii) upgrade  $Q^{\text{UF-A}}$  to  $Q_R^{\text{UF-A}}$  to ensure when an adversary aborts in an unfair manner, then an honest party receives its deposit back plus a predefined amount of compensation plus a reward. Predicates  $Q^{\text{Init}}$  and  $Q^{\text{F-A}}$  remain unchanged. We denote the four predicates as  $\bar{Q} := (Q^{\text{Init}}, Q_R^{\text{Del}}, Q_R^{\text{UF-A}}, Q^{\text{F-A}})$ . We refer readers to Appendix P for formal definitions of predicates  $Q_R^{\text{Del}}$  and  $Q_R^{\text{UF-A}}$ .

Next, we present the formal definition of  $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$ .

**Definition 4** ( $\mathcal{PSI}^{\mathcal{F}^C\mathcal{R}}$ ). *Let  $f^{\text{PSI}}$  be the multi-party PSI functionality defined in Section 4. Protocol  $\Gamma$  realises  $f^{\text{PSI}}$  with  $\bar{Q}$ -fairness-and-reward in the presence of  $m - 3$  static active-adversary clients  $A_j$ s and two rational clients  $A_i$ s or a static passive dealer  $D$  or passive auditor  $Aud$ , if for every non-uniform PPT adversary  $\mathcal{A}$  for the real model, there exists a non-uniform PPT adversary (or simulator)  $\text{Sim}$  for the ideal model, such that for every  $I \in \{A_1, \dots, A_m, D, Aud\}$ , it holds that:  $\{\text{Ideal}_{\text{Sim}(z), I}^{\mathcal{W}(f^{\text{PSI}}, \bar{Q})}(S_1, \dots, S_{m+1})\}_{S_1, \dots, S_{m+1}, z} \stackrel{c}{=} \{\text{Real}_{\mathcal{A}(z), I}^{\Gamma}(S_1, \dots, S_{m+1})\}_{S_1, \dots, S_{m+1}, z}$ , where  $z$  is an auxiliary input given to  $\mathcal{A}$  and  $\mathcal{W}(f^{\text{PSI}}, \bar{Q})$  is a functionality that wraps  $f^{\text{PSI}}$  with predicates  $\bar{Q} := (Q^{\text{Init}}, Q_R^{\text{Del}}, Q_R^{\text{UF-A}}, Q^{\text{F-A}})$ .*

<sup>3</sup> Thus, similar to any  $A_i$  in  $\mathcal{PSI}^{\mathcal{F}^C}$ , these extractors might be corrupted by an active adversary during the PSI computation.

## 8 Concrete Construction of $\mathcal{PSI}^{\mathcal{FCR}}$

### 8.1 Description of Anesidora (ANE)

To construct ANE, we mainly use JUS, deterministic encryption, “double-layered” commitments, the hash-based padding technique, and the counter-collusion smart contracts. At a high level, ANE works as follows. First, all clients run step 1 of JUS to agree on a set of parameters and JUS’s smart contract. They deploy another smart contract, say  $\mathcal{SC}_{\text{ANE}}$ . They also agree on a secret key,  $mk'$ . Next, the buyer places a certain deposit into  $\mathcal{SC}_{\text{ANE}}$ . This deposit will be distributed among honest clients as a reward. The extractors and  $D$  deploy one of the counter-collusion smart contracts, i.e.,  $\mathcal{SC}_{\text{PC}}$ . These three parties deposit a certain amount on this contract. Each honest extractor will receive a portion of  $D$ ’s deposit for carrying out its task honestly and each dishonest extractor will lose a portion of its deposit for acting maliciously. Then, each client encrypts its set elements (under  $mk'$  using deterministic encryption) and then represents the encrypted elements as a polynomial.

Next, the extractors commit to the encryption of their set elements and publish the result. All clients (including  $D$ ) take the rest of the steps in JUS using their input polynomials. This yields a blinded polynomial, whose correctness is checked by  $\mathcal{SC}_{\text{JUS}}$ .

If  $\mathcal{SC}_{\text{JUS}}$  approves the result’s correctness, then all parties receive their deposit put in  $\mathcal{SC}_{\text{JUS}}$ . In this case, each extractor finds the set elements in the intersection. Each extractor proves to  $\mathcal{SC}_{\text{ANE}}$  that the encryptions of the elements in the intersection are among the commitments that the extractor previously published. If  $\mathcal{SC}_{\text{ANE}}$  accepts both extractors’ proofs, then it pays each client (except the buyer) a reward, where the reward is taken from the buyer’s deposit. The extractors receive their deposits back and are paid for carrying out the task honestly. Nevertheless, if  $\mathcal{SC}_{\text{ANE}}$  does not accept one of the extractors’ proofs (or one extractor betrays the other), then it invokes the auditor in the counter-collusion contracts to identify the misbehaving extractor. Then,  $\mathcal{SC}_{\text{ANE}}$  pays each honest client (except the buyer) a reward, taken from the misbehaving extractor.  $\mathcal{SC}_{\text{ANE}}$  also refunds the buyer’s deposit.

If  $\mathcal{SC}_{\text{JUS}}$  does not approve the result’s correctness and  $Aud$  identified misbehaving clients, then honest clients will receive (1) their deposit back from JUS’s contract, and (2) compensation and reward, taken from misbehaving clients. Moreover, the buyer and extractors receive their deposit back from  $\mathcal{SC}_{\text{ANE}}$ . Figure 7 in Appendix Q outlines the interaction between parties in Anesidora.

### 8.2 Detailed Description of ANE

Next, we describe the protocol in more detail (Table 1 in Appendix 3.1 summarises the main notations used).

1. All clients in  $CL = \{A_1, \dots, A_m, D\}$  together run step 1 of JUS (in Section 6.1) to deploy JUS’s contract  $\mathcal{SC}_{\text{JUS}}$  and agree on a master key,  $mk$ .
2. All clients in  $CL$  deploy a new smart contract,  $\mathcal{SC}_{\text{ANE}}$ . The address of  $\mathcal{SC}_{\text{ANE}}$  is given to all clients.

3. Buyer  $A_m$ , before time  $t_1$ , deposits  $S_{min} \cdot \ddot{v}$  amount to  $\mathcal{SC}_{ANE}$ .
4. All clients after time  $t_2 > t_1$  ensure that the buyer has deposited  $S_{min} \cdot \ddot{v}$  amount on  $\mathcal{SC}_{ANE}$ . Otherwise, they abort.
5.  $D$  signs  $\mathcal{SC}_{PC}$  with the extractors.  $\mathcal{SC}_{ANE}$  transfers  $S_{min} \cdot \ddot{r}$  amount (from the buyer deposit) to  $\mathcal{SC}_{PC}$  for each extractor. This is the maximum amount to be paid to an honest extractor for honestly declaring the elements of the intersection. Each extractor deposits  $\ddot{d}' = \ddot{d} + S_{min} \cdot \ddot{f}$  amount in  $\mathcal{SC}_{PC}$  at time  $t_3$ . At time  $t_4$  all clients ensure that the extractors deposited enough coins; otherwise, they withdraw their deposit and abort.
6.  $D$  encrypts  $mk$  under the public key of the dispute resolver (in  $\mathcal{SC}_{PC}$ ); let  $ct_{mk}$  be the resulting ciphertext. It also generates a commitment of  $mk$  as follows:  $z' = \text{PRF}(mk, 0)$ ,  $com_{mk} = \text{Com}(mk, z')$ . It stores  $ct_{mk}$  and  $com_{mk}$  in  $\mathcal{SC}_{ANE}$ .
7. All clients in  $CL$  engage in CT to agree on another key,  $mk'$ .
8. Each client in  $CL$  encrypts the elements of its set  $S : \{s_1, \dots, s_c\}$  as:  $\forall i, 1 \leq i \leq c : e_i = \text{PRP}(mk', s_i)$ . Then, it encodes its encrypted set element as  $\bar{e}_i = e_i || H(e_i)$ . Next, it constructs a hash table HT and inserts the encoded elements into the table.  $\forall i : H(\bar{e}_i) = j$ , then  $\bar{e}_i \rightarrow \text{HT}_j$ . It pads every bin with random dummy elements to  $d$  elements (if needed). Then, for every bin, it builds a polynomial whose roots are the bin's content:  $\pi^{(t)} = \prod_{i=1}^d (x - e'_i)$ , where  $e'_i$  is either  $\bar{e}_i$ , or a dummy value.
9. Every extractor in  $\{A_1, A_2\}$ :
  - (a) for each  $j$ -th bin, commits to the bin's elements:  $com_{i,j} = \text{Com}(e'_i, q_i)$ , where  $q_i$  is a fresh randomness used for the commitment and  $e'_i$  is either  $\bar{e}_i$ , or a dummy value of the bin.
  - (b) constructs a Merkle tree on top of all committed values:  $\text{MT.genTree}(com_{1,1}, \dots, com_{d,h}) \rightarrow g$ .
  - (c) stores the Merkle tree's root  $g$  in  $\mathcal{SC}_{ANE}$ .
10. All clients in  $CL$  run steps 3–11 of JUS, where each client now deposits (in the  $\mathcal{SC}_{JUS}$ )  $\ddot{y}'$  amount where  $\ddot{y}' > S_{min} \cdot \ddot{v} + \ddot{c}h$ . Recall, at the end of step 11 of JUS for each  $j$ -th bin (i) a random polynomial  $\zeta$  is registered in  $\mathcal{SC}_{JUS}$ , (ii) a polynomial  $\phi$  (blinded by a random polynomial  $\gamma'$ ) is extracted by  $\mathcal{SC}_{JUS}$ , and (iii)  $\mathcal{SC}_{JUS}$  checks this polynomial's correctness. If the latter check:
  - passes (i.e.,  $Flag = True$ ), then all parties run step 12 of JUS (with a minor difference, see Appendix T.4). In this case, each party receives  $\ddot{y}'$  amount it deposited in  $\mathcal{SC}_{JUS}$ . They proceed to step 11 below.
  - fails (i.e.,  $Flag = False$ ), then all parties run step 13 of JUS. In this case,  $Aud$  is paid  $\ddot{c}h$  amount, and each honest party receives back its deposit, i.e.,  $\ddot{y}'$  amount. Also, from the misbehaving parties' deposit  $\frac{m' \cdot \ddot{y}' - \ddot{c}h}{m - m'}$  amount is sent to each honest client, to reward and compensate the client  $S_{min} \cdot \ddot{l}$  and  $\frac{m' \cdot \ddot{y}' - \ddot{c}h}{m - m'} - S_{min} \cdot \ddot{l}$  amounts respectively, where  $m'$  is the total number of misbehaving parties. Also,  $\mathcal{SC}_{ANE}$  returns to (a) the buyer its deposit (i.e.,  $S_{min} \cdot \ddot{v}$  amount paid to  $\mathcal{SC}_{ANE}$ ), and (b) each extractor its deposit, i.e.,  $\ddot{d}'$  amount paid to  $\mathcal{SC}_{PC}$ . Then, the protocol halts.
11. Every extractor client:

- (a) finds the elements in the intersection as follows. It encodes each of its set elements to get  $\bar{e}_i$ , as explained in step 8. It determines to which bin the encrypted value belongs, i.e.,  $j = H(\bar{e}_i)$ . It evaluates the resulting polynomial (for that bin) at the encrypted element. It considers the element in the intersection if the evaluation is zero, i.e.,  $\phi(\bar{e}_i) - \zeta(\bar{e}_i) \cdot \gamma'(\bar{e}_i) = 0$ . If the extractor is a traitor, by this step it should have signed  $\mathcal{SC}_{TC}$  with  $D$  and provided related inputs to  $\mathcal{SC}_{TC}$ .
  - (b) proves that every element in the intersection is among the elements it has committed to. Specifically, for each element in the intersection, say  $\bar{e}_i$ , it sends to  $\mathcal{SC}_{ANE}$ :
    - commitment  $com_{i,j}$  (generated in step 9a, for  $\bar{e}_i$ ) and its opening  $\hat{x}' := (\bar{e}_i, q_i)$ .
    - proof  $h_i$  asserting  $com_{i,j}$  is a leaf node of a Merkle tree with root  $g$ .
  - (c) sends the opening of  $com_{mk}$ , i.e., pair  $\hat{x} := (mk, z')$ , to  $\mathcal{SC}_{ANE}$ . This is done only once for all elements in the intersection.
12. Contract  $\mathcal{SC}_{ANE}$ :
- (a) verifies the opening of the commitment for  $mk$ , i.e.,  $\mathbf{Ver}(com_{mk}, \hat{x}) = 1$ . If the verification passes, it generates the index of the bin to which  $\bar{e}_i$  belongs, i.e.,  $j = H(\bar{e}_i)$ . It uses  $mk$  to derive the pseudorandom polynomial  $\gamma'$  for  $j$ -th bin.
  - (b) checks whether (i) the opening of commitment is valid, (ii) the Merkle tree proof is valid, and (iii) the encrypted element is the resulting polynomial's root. Specifically, it ensures that the following relation holds:
 
$$\left( \mathbf{Ver}(com_{i,j}, \hat{x}') = 1 \right) \wedge \left( \mathbf{MT.verify}(h_i, g) = 1 \right) \wedge \left( \phi(\bar{e}_i) - \zeta(\bar{e}_i) \cdot \gamma'(\bar{e}_i) = 0 \right).$$
13. The parties are paid as follows.
- if the extractors' proofs are valid, they provided identical elements of the intersections (for each bin), and there is no traitor, then  $\mathcal{SC}_{ANE}$ :
    - (a) takes  $|S_\cap| \cdot m \cdot \ddot{l}$  amount from the buyer's deposit (in  $\mathcal{SC}_{ANE}$ ) and distributes it among all clients, except the buyer.
    - (b) calls  $\mathcal{SC}_{PC}$  which returns the extractors' deposit (i.e.,  $\ddot{d}'$  amount each) and pays each extractor  $|S_\cap| \cdot \ddot{r}$  amount, for doing their job correctly.
    - (c) checks if  $|S_\cap| < S_{min}$ . If the check passes, then it returns  $(S_{min} - |S_\cap|) \cdot \ddot{v}$  amount to the buyer.
  - if both extractors failed to deliver any result, then  $\mathcal{SC}_{ANE}$ :
    - (a) refunds the buyer, by sending  $S_{min} \cdot \ddot{v}$  amount (deposited in  $\mathcal{SC}_{ANE}$ ) back to the buyer.
    - (b) retrieves each extractor's deposit (i.e.,  $\ddot{d}$  amount) from the  $\mathcal{SC}_{PC}$  and distributes it among the rest of the clients (except the buyer and extractors).
  - Otherwise (e.g., if some proofs are invalid, if an extractor's result is inconsistent with the other extractor's result, or there is a traitor),  $\mathcal{SC}_{ANE}$  invokes (steps 8.c and 9 of)  $\mathcal{SC}_{PC}$  and its auditor to identify the misbehaving extractor, with the help of  $ct_{mk}$  after decrypting it.  $\mathcal{SC}_{ANE}$  asks  $\mathcal{SC}_{PC}$  to pay the auditor the total amount of  $\dot{c}h$  taken from the deposit of the extractor(s) who provided incorrect result to  $\mathcal{SC}_{ANE}$ . Moreover,

- (a) if both extractors cheated:
  - i. if there is no traitor, then  $\mathcal{SC}_{\text{ANE}}$  refunds the buyer, by sending  $S_{\min} \cdot \ddot{v}$  amount (deposited in  $\mathcal{SC}_{\text{ANE}}$ ) back to the buyer. It also distributes  $2 \cdot \ddot{d}' - \ddot{c}h$  amount (taken from the extractors' deposit in  $\mathcal{SC}_{\text{PC}}$ ) among the rest of the clients (except the buyer and extractors).
  - ii. if there is a traitor, then:
    - A. if the traitor delivered a correct result in  $\mathcal{SC}_{\text{TC}}$ ,  $\mathcal{SC}_{\text{ANE}}$  retrieves  $\ddot{d}' - \ddot{d}$  amount from the other dishonest extractor's deposit (in  $\mathcal{SC}_{\text{PC}}$ ) and distributes it among the rest of the clients (except the buyer and dishonest extractor). It asks  $\mathcal{SC}_{\text{PC}}$  to send  $|S_{\cap}| \cdot \ddot{r} + \ddot{d}' + \ddot{d} - \ddot{c}h$  amount to the traitor (via  $\mathcal{SC}_{\text{TC}}$ ).  $\mathcal{SC}_{\text{TC}}$  refunds the traitor's deposit, i.e.,  $\ddot{c}h$  amount. It refunds the buyer, by sending  $S_{\min} \cdot \ddot{v} - |S_{\cap}| \cdot \ddot{r}$  amount (deposited in  $\mathcal{SC}_{\text{ANE}}$ ) to it.
    - B. if the traitor delivered an incorrect result in  $\mathcal{SC}_{\text{TC}}$ , then  $\mathcal{SC}_{\text{ANE}}$  pays the buyer and the rest of the clients in the same way it does in step 13(a)i.  $\mathcal{SC}_{\text{TC}}$  refunds the traitor,  $\ddot{c}h$  amount.
- (b) if one of the extractors cheated:
  - i. if there is no traitor,  $\mathcal{SC}_{\text{ANE}}$  calls  $\mathcal{SC}_{\text{PC}}$  that (a) returns the honest extractor's deposit ( $\ddot{d}'$  amount), (b) pays this extractor  $|S_{\cap}| \cdot \ddot{r}$  amount, for doing its job honestly, and (c) pays this extractor  $\ddot{d} - \ddot{c}h$  amount taken from the dishonest extractor's deposit.  $\mathcal{SC}_{\text{ANE}}$  pays the buyer and the rest of the clients in the same way it does in step 13(a)iiA.
  - ii. if there is a traitor
    - A. if the traitor delivered a correct result in  $\mathcal{SC}_{\text{TC}}$  (but cheated in  $\mathcal{SC}_{\text{ANE}}$ ),  $\mathcal{SC}_{\text{ANE}}$  calls  $\mathcal{SC}_{\text{PC}}$  that (a) returns the other honest extractor's deposit ( $\ddot{d}'$  amount), (b) pays the honest extractor  $|S_{\cap}| \cdot \ddot{r}$  amount taken from the buyer's deposit, for doing its job honestly, (c) pays the honest extractor  $\ddot{d} - \ddot{c}h$  amount taken from the traitor's deposit, (d) pays to the traitor  $|S_{\cap}| \cdot \ddot{r}$  amount taken from the buyer's deposit (via the  $\mathcal{SC}_{\text{TC}}$ ), and (e) refunds the traitor  $\ddot{d}' - \ddot{d}$  amount taken from its own deposit.  $\mathcal{SC}_{\text{TC}}$  refunds the traitor's deposit (i.e.,  $\ddot{c}h$  amount).  $\mathcal{SC}_{\text{ANE}}$  takes  $|S_{\cap}| \cdot m \cdot \ddot{l}$  amount from the buyer's deposit (in  $\mathcal{SC}_{\text{ANE}}$ ) and distributes it among all clients, except the buyer. If  $|S_{\cap}| < S_{\min}$ ,  $\mathcal{SC}_{\text{ANE}}$  returns  $(S_{\min} - |S_{\cap}|) \cdot \ddot{v}$  amount (deposited in  $\mathcal{SC}_{\text{ANE}}$ ) to the buyer.
    - B. if the traitor delivered an incorrect result in  $\mathcal{SC}_{\text{TC}}$  (and it cheated in  $\mathcal{SC}_{\text{ANE}}$ ), then  $\mathcal{SC}_{\text{ANE}}$  pays the honest extractor in the same way it does in step 13(b)iiA.  $\mathcal{SC}_{\text{TC}}$  refunds the traitor's deposit, i.e.,  $\ddot{c}h$  amount. Also,  $\mathcal{SC}_{\text{ANE}}$  pays the buyer and the rest of the clients in the same way it does in step 13(a)iiA.

In Appendix R, we elaborate on the primary challenges we had to confront during the protocol design, which encompassed rewarding clients proportionately to the intersection size and managing collusions among extractors. Additionally,



we direct readers to Appendix T for a more comprehensive discussion of Anesidora, where we elucidate why a naive approach is inadequate and provide insights into the use of hash-based padding and double-layered commitment.

**Theorem 7.** *If PRP, PRF, the commitment scheme, smart contracts, the Merkle tree scheme, JUS and the counter-collusion contracts are secure and the public key encryption is semantically secure, then ANE realises  $f^{PSI}$  with  $\bar{Q}$ -fairness-and-reward (w.r.t. Definition 4) in the presence of  $m - 3$  static active-adversary clients  $A_j$ s and two rational clients  $A_i$ s or a static passive dealer  $D$  or passive auditor  $Aud$ , or passive public which sees the intersection cardinality.*

We prove Theorem 7 in Appendix S.

## 9 Evaluation

In this section, we analyse the asymptotic costs of ANE. We also compare its costs and features with the fastest two and multiple parties PSIs in [2,35,38,42]) and with the fair PSIs in [14,16]. Tables 2 and 3 summarise the result of the cost analysis and the comparison respectively.

Table 2: Asymptotic costs of different parties in ANE. In the table,  $h$  is the total number of bins,  $d$  is a bin’s capacity (i.e.,  $d = 100$ ),  $m$  is the total number of clients (excluding  $D$ ),  $|S|$  is a set cardinality, and  $\bar{\xi}$  is OLE’s security parameter.

Party	Computation Cost	Communication Cost
Client $A_3, \dots, A_m$	$O(h \cdot d(m + d))$	$O(h \cdot d^2 \cdot \bar{\xi})$
Dealer $D$	$O(h \cdot m(d^2 + d))$	$O(h \cdot d^2 \cdot \bar{\xi} \cdot m)$
Auditor $Aud$	$O(h \cdot m \cdot d)$	$O(h \cdot d)$
Extractor $A_1, A_2$	$O(h \cdot d(m + d))$	$O( S  \cdot \log_2  S )$
Smart contract $SC_{ANE}$ & $SC_{JUS}$	$O( S  \cdot (d + \log_2  S ) + h \cdot m \cdot d)$	—
Overall Complexity	$O(h \cdot d^2 \cdot m) = O( S  \cdot d \cdot m)$	$O(h \cdot d^2 \cdot \bar{\xi} \cdot m)$

Table 3: Comparison of the asymptotic complexities and features of state-of-the-art PSIs. In the table,  $t$  is a parameter that determines the maximum number of colluding parties and  $\kappa$  is a security parameter.

Schemes	Asymptotic Cost		Features				
	Computation	Communication	Fairness	Rewarding	Sym-key based	Multi-party	Active Adversary
[2]	$O(h \cdot d^2 \cdot m)$	$O(h \cdot d \cdot m)$	×	×	✓	✓	×
[14]	$O( S )$	$O( S )$	✓	×	×	×	✓
[16]	$O( S ^2)$	$O( S )$	✓	×	×	×	✓
[35]	$O( S  \cdot m^2 +  S  \cdot m)$	$O( S  \cdot m^2)$	×	×	✓	✓	×
[38]	$O( S  \cdot \kappa(m + t^2 - t(m + 1)))$	$O( S  \cdot m \cdot \kappa)$	×	×	✓	✓	✓
[42]	$O( S )$	$O( S  \cdot \kappa)$	×	×	✓	×	✓
<b>Ours: ANE</b>	$O(h \cdot d^2 \cdot m)$	$O(h \cdot d^2 \cdot \bar{\xi} \cdot m)$	✓	✓	✓	✓	✓

### 9.1 Computation Cost

**Client’s and Dealer’s Costs.** In step 1, the cost of each client (including dealer  $D$ ) is  $O(m)$  and mainly involves an invocation of CT. In steps 2–5, the clients’ cost is negligible as it involves deploying smart contracts and reading from them. In step 7, the clients’ cost is  $O(m)$ , as they need to invoke an instance of CT. In step 8, each client invokes PRP and H linear with its set’s cardinality. In the same

step, it also constructs  $h$  polynomials, where the construction of each polynomial involves  $d$  modular multiplications and additions. Thus, its complexity in this step is  $O(h \cdot d)$ . As shown in [2],  $O(h \cdot d) = O(|S|)$  and  $d = 100$  for all set sizes.

In step 10, each client  $A_1, \dots, A_m$  (excluding  $D$ ): (i) invokes an instance of ZSPA-A which involves  $O(h \cdot m)$  invocation of CT,  $3h \cdot m(d+1)$  invocation of PRF,  $3h \cdot m(d+1)$  addition, and  $O(h \cdot m \cdot d)$  invocation of H (in step 3 of subroutine JUS), (ii) invokes  $2h$  instances of VOPR, where each VOPR invocation involves  $2d(1+d)$  invocations of OLE<sup>+</sup>, multiplications, and additions (in steps 6 and 7 of JUS), and (iii) performs  $h(3d+2)$  modular addition (in step 8 of JUS). Also, if  $Flag = True$ , each client (including  $D$ ) invokes PRF  $h(3d+1)$  times, and performs  $h(3d+1)$  additions, and performs polynomial evaluations linear with  $|S|$ , where each evaluation involves  $O(d)$  additions and  $O(d)$  multiplications.

Step 6 involves only  $D$  whose cost in this step is constant, as it involves invoking a public key encryption, PRF, and commitment only once. Furthermore,  $D$ : (a) invokes  $2h \cdot m$  instances of VOPR (in steps 6 and 7 of JUS), (b) invokes PRF  $h(3d+1)$  times (in step 9 of JUS), and (c) performs  $h(d^2+1)$  multiplications and  $3h \cdot m \cdot d$  additions (in step 10 of JUS). If  $Flag = False$ , then  $D$  performs  $O(h \cdot m \cdot d)$  multiplications and additions (in step 13 of JUS).

**Auditor's Cost.** If  $Flag = False$ , then  $Aud$  invokes PRF  $3h \cdot m(d+1)$  times and invokes H  $O(h \cdot m \cdot d)$  times (in step 13 of JUS).

**Extractor's Cost.** In step 9, each extractor invokes the commitment scheme linear with the number of its set cardinality  $|S|$  and constructs a Merkle tree on top of the commitments. In step 11, each extractor invokes H linear with its set cardinality  $|S|$ ; it also performs polynomial evaluations linear with  $|S|$ .

**Smart Contracts' Cost.** In step 10, the subroutine smart contract  $\mathcal{SC}_{JUS}$  performs  $h \cdot m(3d+1)$  additions and  $h$  polynomial divisions, where each division includes dividing a polynomial of degree  $3d+1$  by a polynomial of degree 1 (in step 11 of JUS). In step 12a,  $\mathcal{SC}_{ANE}$  invokes the commitment's verification algorithm **Ver** once, H at most  $|S_\cap|$  times, and PRF  $|S_\cap|(3d+1)$  times. In step 12b,  $\mathcal{SC}_{ANE}$  invokes **Ver** at most  $|S_\cap|$  times, and calls H  $O(|S_\cap| \cdot \log_2 |S|)$  times. In the same step, it performs polynomial evaluation linear with  $|S_\cap|$ . Thus, its overall complexity is  $O(|S_\cap|(d + \log_2 |S|))$ .

## 9.2 Communication Cost

In steps 1 and 7, the communication cost of the clients is dominated by the cost of CT which is  $O(m)$ . In steps 2–6, the clients' cost is negligible, as it involves sending a few transactions to the smart contracts, e.g.,  $\mathcal{SC}_{JUS}$ ,  $\mathcal{SC}_{ANE}$ , and  $\mathcal{SC}_{PC}$ . Step 9 involves only extractors whose cost is  $O(h)$  as each of them only sends to  $\mathcal{SC}_{ANE}$  a single value for each bin. In step 10, the clients' cost is dominated by VOPR's cost; specifically, each pair of clients and  $D$  invokes VOPR  $O(d^2)$  times for each bin; therefore, the cost of each client (excluding  $D$ ) is  $O(h \cdot d^2 \cdot \xi)$  while the cost of  $D$  is  $O(h \cdot d^2 \cdot \xi \cdot m)$ , where  $\xi$  is the subroutine OLE's security parameter. Step 11 involves only the extractors, where each extractor's cost is dominated by the size of the Merkle tree's proof it sends to  $\mathcal{SC}_{ANE}$ , i.e.,  $O(|S_\cap| \cdot \log_2 |S|)$ ,

where  $|S|$  is the extractor’s set cardinality. In step 13, *Aud* sends  $h$  polynomials of degree  $3d + 1$  to  $\mathcal{SC}_{\text{JUS}}$ ; so, its complexity is  $O(h \cdot d)$ . The rest of the steps impose negligible communication costs.

### 9.3 Comparison

Below we demonstrate that ANE provides multiple features not simultaneously offered by state-of-the-art PSIs while maintaining similar overall overhead to efficient PSIs.

**Computation Complexity.** ANE’s computation complexity is similar to that of PSI in [2], but is better than the multiparty PSI’s complexity in [35] as the latter’s complexity is quadratic with the number of parties, i.e.,  $O(|S| \cdot d \cdot m)$  versus  $O(|S| \cdot m^2 + |S| \cdot m)$ . ANE’s complexity is better than the PSI’s complexity in [38] that is quadratic with parameter  $t$ . Similar to the two-party PSIs in [14,42], ANE’s complexity is linear with  $|S|$ . The two-party PSI in [16] imposes a higher computation cost than ANE does, as its complexity is quadratic with sets’ cardinality. Hence, the complexity of ANE is: (i) linear with the set size, similar to the above schemes except the one in [16] and (ii) linear with the total number of parties, similar to the above multi-party schemes, except the one in [35].

**Communication Complexity.** ANE’s communication complexity is slightly higher than the complexity of the PSI in [2], by a factor of  $d \cdot \bar{\xi}$ . However, it is better than the PSI’s complexity in [35] as the latter has a complexity quadratic with the number of parties. ANE’s complexity is slightly higher than the one in [38], by a factor of  $d$ . Similar to the two-party PSIs in [14,42,16], ANE’s complexity is linear with  $c$ . Therefore, the communication complexity of ANE is linear with the set cardinality and number of parties, similar to the above schemes except for the one in [35] whose complexity is quadratic with the number of parties.

**Features.** ANE is the only scheme that offers all the five features, i.e., supports fairness, rewards participants, is based on symmetric key primitives, supports multi-party, and is secure against active adversaries. After ANE is the scheme in [38] which offers three of the above features. The rest of the schemes support only two of the above features. For the sake of fair comparison, we highlight that our ANE and JUS are the only PSIs that use smart contracts (that require additional but standard blockchain-related assumptions), whereas the rest of the above protocols do not use smart contracts.

## 10 Conclusion and Future Direction

PSI stands as a crucial protocol with numerous real-world applications. In this paper, we introduced Justitia, the first multi-party fair PSI, guaranteeing that either all parties obtain the result or, in case of an unfair protocol termination, honest parties receive financial compensation. We then enhanced it into Anesidora, the first PSI that ensures honest parties who contribute their private sets receive rewards proportionate to the number of elements they disclose.

Since the concept of rewarding participants in MPC could potentially boost its real-world adoption, a compelling open question arises: *How can we generalise the idea of rewarding participants in MPC?*

## References

1. Abadi, A., Terzis, S., Metere, R., Dong, C.: Efficient delegated private set intersection on outsourced private datasets. *IEEE TDSC* (2018)
2. Abadi, A., Dong, C., Murdoch, S.J., Terzis, S.: Multi-party updatable delegated private set intersection. In: *FC* (2022)
3. Abadi, A., Dong, C., Murdoch, S.J., Terzis, S.: Multi-party updatable delegated private set intersection –Full Version (2022), <https://fc22.ifca.ai/preproceedings/68.pdf>
4. Abadi, A., Murdoch, S.J., Zacharias, T.: Polynomial representation is tricky: Maliciously secure private set intersection revisited. In: *ESORICS* (2021)
5. Badertscher, C., Gazi, P., Kiayias, A., Russell, A., Zikas, V.: Ouroboros chronos: Permissionless clock synchronization via proof-of-stake. *IACR Cryptol. ePrint Arch.* (2019)
6. Badertscher, C., Maurer, U., Tschudi, D., Zikas, V.: Bitcoin as a transaction ledger: A composable treatment. In: *CRYPTO* (2017)
7. Beimel, A., Omri, E., Orlov, I.: Protocols for multiparty coin toss with dishonest majority. In: *CRYPTO* (2010)
8. Ben-Efraim, A., Nissenbaum, O., Omri, E., Paskin-Cherniavsky, A.: Psimple: Practical multiparty maliciously-secure private set intersection. In: *ASIA CCS* (2022)
9. Berenbrink, P., Czumaj, A., Steger, A., Vöcking, B.: Balanced allocations: the heavily loaded case. In: *STOC* (2000)
10. Blum, M.: Coin flipping by telephone - A protocol for solving impossible problems. In: *COMPCON* (1982)
11. Boneh, D., Gentry, C., Halevi, S., Wang, F., Wu, D.J.: Private database queries using somewhat homomorphic encryption. In: *ACNS* (2013)
12. Cheng, Y., Liu, Y., Chen, T., Yang, Q.: Federated learning for privacy-preserving AI. *Commun. ACM* (2020)
13. Debnath, S.K., Dutta, R.: A fair and efficient mutual private set intersection protocol from a two-way oblivious pseudorandom function. In: *ICISC* (2014)
14. Debnath, S.K., Dutta, R.: New realizations of efficient and secure private set intersection protocols preserving fairness. In: *ICISC* (2016)
15. Debnath, S.K., Dutta, R.: Towards fair mutual private set intersection with linear complexity. *Secur. Commun. Networks* (2016)
16. Dong, C., Chen, L., Camenisch, J., Russello, G.: Fair private set intersection with a semi-trusted arbiter. In: *DBSec*. pp. 128–144 (2013)
17. Dong, C., Wang, Y., Aldweesh, A., McCorry, P., van Moorsel, A.: Betrayal, distrust, and rationality: Smart counter-collusion contracts for verifiable cloud computing. In: *CCS* (2017)
18. Dorn, W.S.: Generalizations of horner’s rule for polynomial evaluation. *IBM Journal of Research and Development* (1962)
19. Freedman, M.J., Nissim, K., Pinkas, B.: Efficient private matching and set intersection. In: *EUROCRYPT* (2004)
20. Garay, J.A., Kiayias, A., Leonardos, N.: The bitcoin backbone protocol: Analysis and applications. In: *Advances in Cryptology - EUROCRYPT* (2015)
21. Ghosh, S., Nielsen, J.B., Nilges, T.: Maliciously secure oblivious linear function evaluation with constant overhead. In: *ASIACRYPT* (2007)
22. Ghosh, S., Nilges, T.: An algebraic approach to maliciously secure private set intersection. In: *EUROCRYPT* (2019)

23. Ghosh, S., Simkin, M.: The communication complexity of threshold private set intersection. In: CRYPTO (2019)
24. Goldreich, O.: The Foundations of Cryptography - Volume 2, Basic Applications. Cambridge University Press (2004)
25. Golob, S., Pentyala, S., Dowsley, R., David, B., Larangeira, M., De Cock, M., Nascimento, A.: A decentralized information marketplace preserving input and output privacy (2023)
26. Inbar, R., Omri, E., Pinkas, B.: Efficient scalable multiparty private set-intersection via garbled bloom filters. In: SCN (2018)
27. Ion, M., Kreuter, B., Nergiz, A.E., Patel, S., Saxena, S., Seth, K., Raykova, M., Shanahan, D., Yung, M.: On deploying secure computing: Private intersection-sum-with-cardinality. In: EuroS&P (2020)
28. Kamara, S., Mohassel, P., Raykova, M., Sadeghian, S.: Scaling private set intersection to billion-element sets. In: FC (2014)
29. Katz, J., Lindell, Y.: Introduction to Modern Cryptography. CRC Press (2007)
30. Kerschbaum, F.: Outsourced private set intersection using homomorphic encryption. In: ASIACCS (2012)
31. Kiayias, A., Russell, A., David, B., Oliynykov, R.: Ouroboros: A provably secure proof-of-stake blockchain protocol. In: CRYPTO (2017)
32. Kiayias, A., Zhou, H., Zikas, V.: Fair and robust multi-party computation using a global transaction ledger. In: EUROCRYPT (2016)
33. Kissner, L., Song, D.X.: Privacy-preserving set operations. In: CRYPTO (2005)
34. Koch, K., Krenn, S., Marc, T., More, S., Ramacher, S.: KRAKEN: a privacy-preserving data market for authentic data. In: DE. ACM (2022)
35. Kolesnikov, V., Matania, N., Pinkas, B., Rosulek, M., Trieu, N.: Practical multiparty private set intersection from symmetric-key techniques. In: CCS (2017)
36. Moran, T., Naor, M., Segev, G.: An optimally fair coin toss. In: TCC (2009)
37. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. Tech. rep. (2019)
38. Nevo, O., Trieu, N., Yanai, A.: Simple, fast malicious multiparty private set intersection. In: CCS (2021)
39. Papamanthou, C., Tamassia, R., Triandopoulos, N.: Authenticated hash tables. In: CCS (2008)
40. Pedersen, T.P.: Non-interactive and information-theoretic secure verifiable secret sharing. In: CRYPTO (1991)
41. Pinkas, B., Schneider, T., Segev, G., Zohner, M.: Phasing: Private set intersection using permutation-based hashing. In: USENIX Security (2015)
42. Raghuraman, S., Rindal, P.: Blazing fast PSI from improved OKVS and subfield VOLE. In: CCS (2022)
43. Schneier, B.: Applied cryptography - protocols, algorithms, and source code in C, 2nd Edition. Wiley (1996)
44. Tamrakar, S., Liu, J., Paverd, A., Ekberg, J., Pinkas, B., Asokan, N.: The circle game: Scalable private membership test using trusted hardware. In: AsiaCCS (2017)
45. Tian, Z., Liu, J., Li, J., Cao, X., Jia, R., Ren, K.: Private data valuation and fair payment in data marketplaces. CoRR (2022)
46. Wood, G., et al.: Ethereum: A secure decentralised generalised transaction ledger. Ethereum project yellow paper (2014)
47. Zhang, E., Liu, F., Lai, Q., Jin, G., Li, Y.: Efficient multi-party private set intersection against malicious adversaries. In: CCSW (2019)
48. Zhao, Y., Chow, S.S.M.: Can you find the one for me? privacy-preserving match-making via threshold PSI. ePrint Arch. (2018)

## A Full Security Model

In this paper, we use the simulation-based paradigm of secure computation [24] to define and prove the proposed protocols. Since both types of (static) active and passive adversaries are involved in our protocols, we will provide formal definitions for both types.

**Two-party Computation.** A two-party protocol  $\Gamma$  problem is captured by specifying a random process that maps pairs of inputs to pairs of outputs, one for each party. Such process is referred to as a functionality denoted by  $f : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^* \times \{0, 1\}^*$ , where  $f := (f_1, f_2)$ . For every input pair  $(x, y)$ , the output pair is a random variable  $(f_1(x, y), f_2(x, y))$ , such that the party with input  $x$  wishes to obtain  $f_1(x, y)$  while the party with input  $y$  wishes to receive  $f_2(x, y)$ . When  $f$  is deterministic, then  $f_1 = f_2$ . In the setting where  $f$  is asymmetric and only one party (say the first one) receives the result,  $f$  is defined as  $f := (f_1(x, y), \perp)$ .

**Security in the Presence of Passive Adversaries.** In the passive adversarial model, the party corrupted by such an adversary correctly follows the protocol specification. Nonetheless, the adversary obtains the internal state of the corrupted party, including the transcript of all the messages received, and tries to use this to learn information that should remain private. Loosely speaking, a protocol is secure if whatever can be computed by a party in the protocol can be computed using its input and output only. In the simulation-based model, it is required that a party's view in a protocol's execution can be simulated given only its input and output. This implies that the parties learn nothing from the protocol's execution. More formally, party  $i$ 's view (during the execution of  $\Gamma$ ) on input pair  $(x, y)$  is denoted by  $\text{View}_i^\Gamma(x, y)$  and equals  $(w, r^i, m_1^i, \dots, m_t^i)$ , where  $w \in \{x, y\}$  is the input of  $i^{\text{th}}$  party,  $r_i$  is the outcome of this party's internal random coin tosses, and  $m_j^i$  represents the  $j^{\text{th}}$  message this party receives. The output of the  $i^{\text{th}}$  party during the execution of  $\Gamma$  on  $(x, y)$  is denoted by  $\text{Output}_i^\Gamma(x, y)$  and can be generated from its own view of the execution. The joint output of both parties is denoted by  $\text{Output}^\Gamma(x, y) := (\text{Output}_1^\Gamma(x, y), \text{Output}_2^\Gamma(x, y))$ .

**Definition 5.** Let  $f$  be the deterministic functionality defined above. Protocol  $\Gamma$  security computes  $f$  in the presence of a static passive adversary if there exist polynomial-time algorithms  $(\text{Sim}_1, \text{Sim}_2)$  such that:

$$\begin{aligned} \{\text{Sim}_1(x, f_1(x, y))\}_{x, y} &\stackrel{c}{=} \{\text{View}_1^\Gamma(x, y)\}_{x, y} \\ \{\text{Sim}_2(x, f_2(x, y))\}_{x, y} &\stackrel{c}{=} \{\text{View}_2^\Gamma(x, y)\}_{x, y} \end{aligned}$$

**Security in the Presence of Active Adversaries.** In this adversarial model, the corrupted party may arbitrarily deviate from the protocol specification, to learn the private inputs of the other parties or to influence the outcome of the computation. In this case, the adversary may not use the input provided.

Therefore, beyond the possibility that a corrupted party may learn more than it should, correctness is also required. This means that a corrupted party must not be able to cause the output to be incorrectly distributed. Moreover, we require independence of inputs meaning that a corrupted party cannot make its input depend on the other party's input. To capture the threats, the security of a protocol is analyzed by comparing what an adversary can do in the real protocol to what it can do in an ideal scenario that is secure by definition. This is formalized by considering an ideal computation involving an incorruptible Trusted Third Party (TTP) to whom the parties send their inputs and receive the output of the ideal functionality. Below, we describe the executions in the ideal and real models.

First, we describe the execution in the ideal model. Let  $P_1$  and  $P_2$  be the parties participating in the protocol,  $i \in \{0, 1\}$  be the index of the corrupted party, and  $\mathcal{A}$  be a non-uniform probabilistic polynomial-time adversary. Also, let  $z$  be an auxiliary input given to  $\mathcal{A}$  while  $x$  and  $y$  be the input of party  $P_1$  and  $P_2$  respectively. The honest party,  $P_j$ , sends its received input to TTP. The corrupted party  $P_i$  may either abort (by replacing the input with a special abort message  $A_i$ ), send its received input or send some other input of the same length to TTP. This decision is made by the adversary and may depend on the input value of  $P_i$  and  $z$ . If TTP receives  $A_i$ , then it sends  $A_i$  to the honest party and the ideal execution terminates. Upon obtaining an input pair  $(x, y)$ , TTP computes  $f_1(x, y)$  and  $f_2(x, y)$ . It first sends  $f_i(x, y)$  to  $P_i$  which replies with "continue" or  $A_i$ . In the former case, TTP sends  $f_j(x, y)$  to  $P_j$  and in the latter it sends  $A_i$  to  $P_j$ . The honest party always outputs the message that it obtained from TTP. A malicious party may output an arbitrary function of its initial inputs and the message it has obtained from TTP. The ideal execution of  $f$  on inputs  $(x, y)$  and  $z$  is denoted by  $\text{Ideal}_{\mathcal{A}(z), i}^f(x, y)$  and is defined as the output pair of the honest party and  $\mathcal{A}$  from the above ideal execution. In the real model, the real two-party protocol  $\Gamma$  is executed without the involvement of TTP. In this setting,  $\mathcal{A}$  sends all messages on behalf of the corrupted party and may follow an arbitrary strategy. The honest party follows the instructions of  $\Gamma$ . The real execution of  $\Gamma$  is denoted by  $\text{Real}_{\mathcal{A}(z), i}^\Gamma(x, y)$ , it is defined as the joint output of the parties engaging in the real execution of  $\Gamma$  (on the inputs), in the presence of  $\mathcal{A}$ .

Next, we define security. At a high level, the definition states that a secure protocol in the real model emulates the ideal model. This is formulated by stating that adversaries in the ideal model can simulate executions of the protocol in the real model.

**Definition 6.** *Let  $f$  be the two-party functionality defined above and  $\Gamma$  be a two-party protocol that computes  $f$ . Protocol  $\Gamma$  securely computes  $f$  with abort in the presence of static active adversaries if for every non-uniform probabilistic polynomial time adversary  $\mathcal{A}$  for the real model, there exists a non-uniform probabilistic polynomial-time adversary (or simulator)  $\text{Sim}$  for the ideal model, such that for every  $i \in \{0, 1\}$ , it holds that:*

$$\{\text{Ideal}_{\text{Sim}(z), i}^f(x, y)\}_{x, y, z} \stackrel{c}{=} \{\text{Real}_{\mathcal{A}(z), i}^\Gamma(x, y)\}_{x, y, z}$$

## B Counter Collusion Contracts

In this section, we present Prisoner’s Contract ( $\mathcal{SC}_{\text{PC}}$ ), Colluder’s Contract ( $\mathcal{SC}_{\text{CC}}$ ), Traitor’s Contract ( $\mathcal{SC}_{\text{TC}}$ ) originally proposed by Dong *et al.* [17]. As we previously stated, we have slightly adjusted the contracts. We will highlight the adjustments in blue. For the sake of completeness, below we restate the parameters used in these contracts and their relation.

- $\ddot{b}$ : the bribe paid by the ringleader of the collusion to the other server in the collusion agreement, in  $\mathcal{SC}_{\text{CC}}$ .
- $\ddot{c}$ : a server’s cost for computing the task.
- $\dot{c}h$ : the fee paid to invoke an auditor for recomputing a task and resolving disputes.
- $\ddot{d}$ : the deposit a server needs to pay to be eligible for getting the job.
- $\ddot{t}$ : the deposit the colluding parties need to pay in the collusion agreement, in  $\mathcal{SC}_{\text{CC}}$ .
- $\ddot{w}$ : the amount that a server receives for completing the task.
- $\ddot{w} \geq \ddot{c}$ : the server would not accept underpaid jobs.
- $\dot{c}h > 2\ddot{w}$ : if it does not hold, then there would be no need to use the servers and the auditor would do the computation.
- $(pk, sk)$ : an asymmetric-key encryption’s public-private key pair belonging to the auditor.

The following relations need to hold when setting the contracts in order for the desirable equilibria to hold: (i)  $\ddot{d} > \ddot{c} + \dot{c}h$ , (ii)  $\ddot{b} < \ddot{c}$ , and (iii)  $\ddot{t} < \ddot{w} - \ddot{c} + 2\ddot{d} - \dot{c}h - \ddot{b}$ .

### B.1 Prisoner’s Contract ( $\mathcal{SC}_{\text{PC}}$ )

$\mathcal{SC}_{\text{PC}}$  has been designed for outsourcing a certain computation. It is signed by a client who delegates the computation and two other parties (or servers) who perform the computation. This contract tries to incentivize correct computation by using the following idea. It requires each server to pay a deposit before the computation is delegated. If a server behaves honestly, then it can withdraw its deposit. However, if a server cheats (and is detected), its deposit is transferred to the client. When one of the servers is honest and the other one cheats, the honest server receives a reward. This reward is taken from the deposit of the cheating server. Hence, the goal of  $\mathcal{SC}_{\text{PC}}$  is to create a Prisoner’s dilemma between the two servers in the following sense. Although the servers may collude with each other (to cut costs and provide identical but incorrect computation results) which leads to a higher payoff than both behaving honestly, there is an even higher payoff if one of the servers manages to persuade the other server to collude and provide an incorrect result while itself provides a correct result. In this setting, each server knows that collusion is not stable as its counterparty will always try to deviate from the collusion to increase its payoff. If a server tries to convince its counterparty (without a credible and enforceable promise), then the latter party will consider it as a trap; consequently, collusion will not occur. Below, we restate the contract.



1. The contract is signed by three parties; namely, client  $D$  and two other parties  $E_1$  and  $E_2$ . A third-party auditor will resolve any dispute between  $D$  and the servers. The address of another contract, called  $\mathcal{SC}_{\text{ANE}}$ , is hardcoded in this contract.
2. The servers agree to run computation  $f$  on input  $x$ , both of which have been provided by  $D$ .
3. The parties agree on three deadlines  $(T_1, T_2, T_3)$ , where  $T_{i+1} > T_i$ .
4.  $D$  agrees to pay  $\ddot{w}$  to each server for the correct and on-time computation. Therefore,  $D$  deposits  $2 \cdot \ddot{w}$  amount in the contract. This deposit is transferred from  $\mathcal{SC}_{\text{ANE}}$  to this contract.
5. Each server deposits  $\ddot{d}' = \ddot{d} + \ddot{X}$  amount in the contract.
6. The servers must pay the deposit before  $T_1$ . If a server fails to meet the deadline, then the contract would refund the parties' deposit (if any) and terminates.
7. The servers must deliver the computation's result before  $T_2$ .
8. The following steps are taken when (i) both servers provided the computation's result or (ii) deadline  $T_2$  elapsed.
  - (a) if both servers failed to deliver the computation's result, then the contract transfers their deposits to  $\mathcal{SC}_{\text{ANE}}$ .
  - (b) if both servers delivered the result, and the results are equal, then (after verifying the results) this contract must pay the agreed amount  $\ddot{w}$  and refund the deposit  $\ddot{d}'$  to each server.
  - (c) otherwise,  $D$  raises a dispute with the auditor.
9. When a dispute is raised, the auditor (which is independent of  $Aud$  in  $JUS$ ) re-generates the computation's result, by using algorithm  $\text{resComp}(\cdot)$  described shortly in Appendix B.1. Let  $y_i, y_1$ , and  $y_2$  be the result computed by the auditor,  $E_1$ , and  $E_2$  respectively. The auditor uses the following role to identify the cheating party.
  - if  $E_i$  failed to deliver the result (i.e.,  $y_i$  is null), then it has cheated.
  - if a result  $y_i$  has been delivered before the deadline and  $y_i \neq y_i$ , then  $E_i$  has cheated.
 The auditor sends its verdict to  $\mathcal{SC}_{\text{PC}}$ .
10. Given the auditor's decision, the dispute is settled according to the following rules.
  - if none of the servers cheated, then this contract transfers to each server (i)  $\ddot{w}$  amount for performing the computation and (ii) its deposit, i.e.,  $\ddot{d}'$  amount. The client also pays the auditor  $\check{c}\check{h}$  amount.
  - if both servers cheated, then this contract (i) pays the auditor the total amount of  $\check{c}\check{h}$ , taken from the servers' deposit, and (ii) transfers to  $\mathcal{SC}_{\text{ANE}}$  the rest of the deposit, i.e.,  $2 \cdot \ddot{d}' - \check{c}\check{h}$  amount.
  - if one of the servers cheated, then this contract (i) pays the auditor the total amount of  $\check{c}\check{h}$ , taken from the misbehaving server's deposit, (ii) transfers the honest server's deposit (i.e.,  $\ddot{d}'$  amount) back to this server, (iii) transfers to the honest server  $\ddot{w} + \ddot{d} - \check{c}\check{h}$  amount (which covers its computation cost and the reward), and (iv) transfers to  $\mathcal{SC}_{\text{ANE}}$  the rest of the misbehaving server's deposit, i.e.,  $\ddot{X}$  amount. The cheating server receives nothing.

11. After deadline  $T_3$ , if  $D$  has neither paid nor raised a dispute, then [this contract](#) pays  $\ddot{w}$  to each server which delivered a result before deadline  $T_2$  and refunds each server its deposit, i.e.,  $\ddot{d}'$  amount. Any deposit left after that will be transferred to  $\mathcal{SC}_{\text{ANE}}$ .

Now, we explain why we have made the above changes to the original  $\mathcal{SC}_{\text{PC}}$  of Dong *et al.* [17]. In the original  $\mathcal{SC}_{\text{PC}}$  (a) the client does not deposit any amount in this contract; instead, it directly sends its coins to a server (and auditor) according to the auditor's decision, (b) the computation correctness is determined only within this contract (with the involvement of the auditor if required), and (c) the auditor simply re-generates the computation's result given the computation's plaintext inputs. Nevertheless, in ANE, (1) *all clients* need to deposit a certain amount in  $\mathcal{SC}_{\text{ANE}}$  and only the contracts must transfer the parties' deposit, (2)  $\mathcal{SC}_{\text{ANE}}$  also needs to verify a part of the computation's correctness without the involvement of the auditor and accordingly distribute the parties' deposit based on the verification's outcome, and (3) the auditor must be able to re-generate the computation's result without being able to learn the computation's plaintext input, i.e., elements of the set. Therefore, we have included the address of  $\mathcal{SC}_{\text{ANE}}$  in  $\mathcal{SC}_{\text{PC}}$  to let the parties' deposit move between the two contracts (if necessary) and have allowed  $\mathcal{SC}_{\text{PC}}$  to distribute the parties' deposit; thus, the requirements in points (1) and (2) are met. To meet the requirement in point (3) above, we have included a new algorithm, called  $\text{resComp}(\cdot)$ , in  $\mathcal{SC}_{\text{PC}}$ . Shortly, we will provide more detail about this algorithm. Moreover, to make this contract compatible with ANE, we increased the amount of each server's deposit by  $\ddot{X}$ . Nevertheless, this adjustment does not change the logic behind  $\mathcal{SC}_{\text{PC}}$ 's design and its analysis.

**Auditor's Result-Computation Algorithm.** In this work, we use  $\mathcal{SC}_{\text{PC}}$  to delegate the computation of intersection cardinality to two extractor clients, a.k.a. servers in the original  $\mathcal{SC}_{\text{PC}}$ . In this setting, the contract's auditor is invoked when an inconsistency is detected in step 13 of ANE. For the auditor to recompute the intersection cardinality, we have designed  $\text{resComp}(\cdot)$  algorithm. The auditor uses this algorithm for every bin's index  $\text{indx}$ , where  $1 \leq \text{indx} \leq h$  and  $h$  is the hash table's length. We present this algorithm in Figure 4. The auditor collects the inputs of this algorithm as follows: (a) reads random polynomial  $\zeta$ , and blinded polynomial  $\phi$  from contract  $\mathcal{SC}_{\text{JUS}}$ , (b) reads the ciphertext if secret key  $mk$  from  $\mathcal{SC}_{\text{ANE}}$ , and (c) fetches public parameters  $(\text{des}_H, h)$  from the hash table's public description.

Note that in the original  $\mathcal{SC}_{\text{PC}}$  of Dong *et al.* [17], the auditor is assumed to be fully trusted. However, in this work, we have relaxed such an assumption. We have designed ANE and  $\text{resComp}(\cdot)$  in such a way that even a semi-honest auditor cannot learn anything about the actual elements of the sets, as they have been encrypted under a key unknown to the auditor.

$\text{resComp}(\zeta, \phi, sk, ct_{mk}, indx, des_H) \rightarrow R$

- *Input.*  $\zeta$ : a random polynomial of degree 1,  $\phi$ : a blinded polynomial of the form  $\zeta \cdot (\epsilon + \gamma')$  where  $\epsilon$  and  $\gamma'$  are arbitrary and pseudorandom polynomials respectively,  $\deg(\phi) - 1 = \deg(\gamma')$ ,  $sk$ : the auditor's secret key,  $ct_{mk}$ : ciphertext of  $mk$  which is a key of PRF,  $indx$ : an input of PRF, and  $des_H$ : a description of hash function  $H$ .
  - *Output.*  $R$ : a set containing valid roots of unblinded  $\phi$ .
1. decrypts the ciphertext  $ct_{mk}$  under key  $sk$ . Let  $mk$  be the result.
  2. unblinds polynomial  $\phi$ , as follows:
    - (a) re-generates pseudorandom polynomial  $\gamma'$  using key  $mk$ . Specifically, it uses  $mk$  to derive a key:  $k = \text{PRF}(mk, indx)$ . Then, it uses the derived key to generate  $3d+1$  pseudorandom coefficients, i.e.,  $\forall j, 0 \leq j \leq \deg(\phi) - 1 : g_j = \text{PRF}(k, j)$ . Next, it uses these coefficients to construct polynomial  $\gamma'$ , i.e.,  $\gamma' = \sum_{j=0}^{\deg(\phi)-1} g_j \cdot x^j$ .
    - (b) removes the blinding factor from  $\phi$ . Specifically, it computes polynomial  $\phi'$  of the following form  $\phi' = \phi - \zeta \cdot \gamma'$ .
  3. extracts roots of polynomial  $\phi'$ .
  4. finds valid roots, by (i) parsing each root  $\bar{e}$  as  $(e_1, e_2)$  with the assistance of  $des_H$  and (ii) checking if  $e_2 = H(e_1)$ . It considers a root valid, if this equation holds.
  5. returns set  $R$  containing all valid roots.

Fig. 4: Auditor's result computation,  $\text{resComp}(\cdot)$ , algorithm

## B.2 Colluder's Contract ( $\mathcal{SC}_{CC}$ )

Recall that  $\mathcal{SC}_{PC}$  aimed at creating a dilemma between the two servers. However, this dilemma can be addressed if they can make an enforceable promise. This enforceable promise can be another smart contract, called Colluder's Contract ( $\mathcal{SC}_{CC}$ ). This contract imposes additional rules that ultimately would affect the parties' payoffs and would make collusion the most profitable strategy for the colluding parties. In  $\mathcal{SC}_{CC}$ , the party who initiates the collusion would pay its counterparty a certain amount (or bribe) if both follow the collusion and provide an incorrect result of the computation to  $\mathcal{SC}_{PC}$ . Note,  $\mathcal{SC}_{CC}$  requires both servers to send a fixed amount of deposit when signing the contract. The party who deviates from collusion will be punished by losing the deposit. Below, we restate  $\mathcal{SC}_{CC}$ .

1. The contract is signed between the server who initiates the collusion, called ringleader (LDR) and the other server called follower (FLR).
2. The two agree on providing to  $\mathcal{SC}_{PC}$  a different result  $res'$  than a correct computation of  $f$  on  $x$  would yield, i.e.,  $res' \neq f(x)$ . Parameter  $res'$  is recorded in this contract.
3. LDR and FLR deposit  $\tilde{t} + \tilde{b}$  and  $\tilde{t}$  amounts in this contract respectively.

4. The above deposit must be paid before the result delivery deadline in  $\mathcal{SC}_{PC}$ , i.e., before deadline  $T_2$ . If this condition is not met, the parties' deposit in this contract is refunded and this contract is terminated.
5. When  $\mathcal{SC}_{PC}$  is finalised (i.e., all the results have been provided), the following steps are taken.
  - (a) both follow the collusion: if both LDR and FLR provided  $res'$  to  $\mathcal{SC}_{PC}$ , then  $\ddot{t}$  and  $\ddot{t} + \ddot{b}$  amounts are delivered to LDR and FLR respectively. Therefore, FLR receives its deposit plus the bribe  $\ddot{b}$ .
  - (b) only FLR deviates from the collusion: if LDR and FLR provide  $res'$  and  $res'' \neq res'$  to  $\mathcal{SC}_{PC}$  respectively, then  $2 \cdot \ddot{t} + \ddot{b}$  amount is transferred to LDR while nothing is sent to FLR.
  - (c) only LDR deviates from the collusion: if LDR and FLR provide  $res'' \neq res'$  and  $res'$  to  $\mathcal{SC}_{PC}$  respectively, then  $2 \cdot \ddot{t} + \ddot{b}$  amount is sent to FLR while nothing is transferred to LDR.
  - (d) both deviate from the collusion: if LDR and FLR deviate and provide any result other than  $res'$  to  $\mathcal{SC}_{PC}$ , then  $2 \cdot \ddot{t} + \ddot{b}$  amount is sent to LDR and  $\ddot{t}$  amount is sent to FLR.

We highlight that the amount of bribe a rational LDR is willing to pay is less than its computation cost (i.e.,  $\ddot{b} < \ddot{c}$ ); otherwise, such collusion would not bring a higher payoff. We refer readers to [17] for further discussion.

### B.3 Traitor's Contract ( $\mathcal{SC}_{TC}$ )

$\mathcal{SC}_{TC}$  incentivises a colluding server (who has had signed  $\mathcal{SC}_{CC}$ ) to betray its counterparty and report the collusion without being penalised by  $\mathcal{SC}_{PC}$ . The Traitor's contract promises that the reporting server will not be punished by  $\mathcal{SC}_{PC}$  which makes it safe for the reporting server to follow the collusion strategy (of  $\mathcal{SC}_{CC}$ ), and get away from the punishment imposed by  $\mathcal{SC}_{PC}$ . Below, we restate  $\mathcal{SC}_{TC}$ .

1. This contract is signed among  $D$  and the traitor server (TRA) who reports the collusion. This contract is signed only if the parties have already signed  $\mathcal{SC}_{PC}$ .
2.  $D$  signs this contract only with the first server who reports the collusion.
3. The traitor TRA must also provide to this contract the result of the computation, i.e.,  $f(x)$ . The result provided in this contract could be different than the one provided in  $\mathcal{SC}_{PC}$ , e.g., when TRA has to follow  $\mathcal{SC}_{CC}$  and provide an incorrect result to  $\mathcal{SC}_{PC}$ .
4.  $D$  needs to pay  $\ddot{w} + \ddot{d}' + \ddot{d} - \ddot{c}h$  amount to this contract. This amount equals the maximum amount TRA could lose in  $\mathcal{SC}_{PC}$  plus the reward. **This deposit will be transferred via  $\mathcal{SC}_{ANE}$  to this contract.** TRA must deposit in this contract  $\ddot{c}h$  amount to cover the cost of resolving a potential dispute.
5. This contract should be signed before the deadline  $T_2$  for the delivery of the computation result in  $\mathcal{SC}_{PC}$ . If it is not signed on time, then this contract would be terminated and any deposit paid will be refunded.

6. It is required that the TRA provide the computation result to this contract before the above deadline  $T_2$ .
7. If this contract is fully signed, then during the execution of  $\mathcal{SC}_{\text{PC}}$ ,  $D$  always raises a dispute, i.e., takes step 8c in  $\mathcal{SC}_{\text{PC}}$ .
8. After  $\mathcal{SC}_{\text{PC}}$  is finalised (with the involvement of the auditor), the following steps are taken to pay the parties involved.
  - (a) if none of the servers cheated in  $\mathcal{SC}_{\text{PC}}$  (according to the auditor), then amount  $\ddot{w} + \ddot{d}' + \ddot{d} - \ddot{c}h$  is refunded to  $\mathcal{SC}_{\text{ANE}}$  and TRA's deposit (i.e.,  $\ddot{c}h$  amount) is transferred to  $D$ . Nothing is paid to TRA.
  - (b) if in  $\mathcal{SC}_{\text{PC}}$ , the other server did not cheat and TRA cheated; however, TRA provided a correct result in this contract, then  $\ddot{d}' + \ddot{d} - \ddot{c}h$  amount is transferred to  $\mathcal{SC}_{\text{ANE}}$ . Also, TRA gets its deposit back (i.e.,  $\ddot{c}h$  amount) plus  $\ddot{w}$  amounts for providing a correct result to this contract.
  - (c) if in  $\mathcal{SC}_{\text{PC}}$ , both servers cheated; however, TRA delivered a correct computation result to this contract, then TRA gets its deposit back (i.e.,  $\ddot{c}h$  amount), it also receives  $\ddot{w} + \ddot{d}' + \ddot{d} - \ddot{c}h$  amount.
  - (d) otherwise,  $\ddot{w} + \ddot{d}' + \ddot{d} - \ddot{c}h$  and  $\ddot{c}h$  amounts are transferred to  $\mathcal{SC}_{\text{ANE}}$  and TRA respectively.
9. If TRA provided a result to this contract, and deadline  $T_3$  (in  $\mathcal{SC}_{\text{PC}}$ ) has passed, then all deposits, if any left, will be transferred to TRA.

TRA must take the following three steps to report collusion: (i) it waits until  $\mathcal{SC}_{\text{CC}}$  is signed by the other server, (ii) it reports the collusion to  $D$  before signing  $\mathcal{SC}_{\text{CC}}$ , and (iii) it signs  $\mathcal{SC}_{\text{CC}}$  only after it signed  $\mathcal{SC}_{\text{TC}}$  with  $D$ . Note, the original analysis of  $\mathcal{SC}_{\text{TC}}$  does not require  $\mathcal{SC}_{\text{TC}}$  to remain secret. Therefore, in our ANE, parties TRA and  $D$  can sign this contract and then store its address in  $\mathcal{SC}_{\text{ANE}}$ . Alternatively, to keep this contract confidential,  $D$  can deploy a template  $\mathcal{SC}_{\text{TC}}$  to the blockchain and store the commitment of the contract's address (instead of the plaintext address) in  $\mathcal{SC}_{\text{ANE}}$ . When a traitor (TRA) wants to report collusion, it signs the deployed  $\mathcal{SC}_{\text{TC}}$  with  $D$  which provides the commitment opening to TRA. In this case, at the time when the deposit is distributed, either  $D$  or TRA provides the opening of the commitment to  $\mathcal{SC}_{\text{ANE}}$  which checks whether it matches the commitment. If the check passes, then it distributes the deposit as before.

## C Commitment Scheme

A commitment scheme involves a *sender* and a *receiver*. It also involves two phases; namely, *commit* and *open*. In the commit phase, the sender commits to a message:  $x$  as  $\text{Com}(x, r) = \text{com}$ , that involves a secret value:  $r \xleftarrow{\$} \{0, 1\}^\lambda$ . At the end of the commit phase, the commitment  $\text{com}$  is sent to the receiver. In the open phase, the sender sends the opening  $\hat{x} := (x, r)$  to the receiver who verifies its correctness:  $\text{Ver}(\text{com}, \hat{x}) \stackrel{?}{=} 1$  and accepts if the output is 1. A commitment scheme must satisfy two properties: (a) *hiding*: it is infeasible for an adversary (i.e., the receiver) to learn any information about the committed

message  $x$ , until the commitment  $com$  is opened, and (b) *binding*: it is infeasible for an adversary (i.e., the sender) to open a commitment  $com$  to different values  $\hat{x}' := (x', r')$  than that was used in the commit phase, i.e., infeasible to find  $\hat{x}'$ , s.t.  $\mathbf{Ver}(com, \hat{x}) = \mathbf{Ver}(com, \hat{x}') = 1$ , where  $\hat{x} \neq \hat{x}'$ . There exist efficient commitment schemes both in (a) the standard model, e.g., Pedersen scheme [40], and (b) the random oracle model using the well-known hash-based scheme such that committing is  $H(x||r) = com$  and  $\mathbf{Ver}(com, \hat{x})$  requires checking:  $H(x||r) \stackrel{?}{=} com$ , where  $H: \{0, 1\}^* \rightarrow \{0, 1\}^\lambda$  is a collision-resistant hash function, i.e., the probability to find  $x$  and  $x'$  such that  $H(x) = H(x')$  is negligible in the security parameter  $\lambda$ .

## D Hash Tables

We set the table's parameters appropriately to ensure the number of elements in each bin does not exceed a predefined capacity. Given the maximum number of elements  $c$  and the bin's maximum size  $d$ , we can determine the number of bins by analysing hash tables under the balls into bins model [9].

**Theorem 8. (Upper Tail in Chernoff Bounds)** *Let  $X_i$  be a random variable defined as  $X_i = \sum_{j=1}^c Y_{ij}$ , where  $Pr[Y_{ij} = 1] = p_i$ ,  $Pr[Y_{ij} = 0] = 1 - p_i$ , and all  $Y_{ij}$  are independent. Let the expectation be  $\mu = E[X_i] = \sum_{j=1}^c p_i$ , then  $Pr[X_i > d = (1 + \sigma) \cdot \mu] < \left( \frac{e^\sigma}{(1 + \sigma)^{(1 + \sigma)}} \right)^\mu, \forall \sigma > 0$*

In this model, the expectation is  $\mu = \frac{c}{h}$ , where  $c$  is the number of balls and  $h$  is the number of bins. The above inequality provides the probability that bin  $i$  gets more than  $(1 + \sigma) \cdot \mu$  balls. Since there are  $h$  bins, the probability that at least one of them is overloaded is bounded by the union bound:

$$Pr[\exists i, X_i > d] \leq \sum_{i=1}^h Pr[X_i > d] = h \cdot \left( \frac{e^\sigma}{(1 + \sigma)^{(1 + \sigma)}} \right)^{\frac{c}{h}} \quad (1)$$

Thus, for a hash table of length  $h = O(c)$ , there is always an *almost constant* expected number of elements,  $d$ , mapped to the same bin with a high probability [39], e.g.,  $1 - 2^{-40}$ .

Appendix N.5 provides an explanation of how to establish concrete parameters for a hash table.

## E Merkle Tree

A Merkle tree is a data structure that supports a compact commitment of a set of values/blocks. As a result, it includes two parties, prover  $\mathcal{P}$  and verifier  $\mathcal{V}$ . The Merkle tree scheme includes three algorithms ( $\mathbf{MT.genTree}$ ,  $\mathbf{MT.prove}$ ,  $\mathbf{MT.verify}$ ), defined as follows:

- The algorithm that constructs a Merkle tree, **MT.genTree**, is run by  $\mathcal{V}$ . It takes blocks,  $u := u_1, \dots, u_n$ , as input. Then, it groups the blocks in pairs. Next, a collision-resistant hash function,  $H(\cdot)$ , is used to hash each pair. After that, the hash values are grouped in pairs and each pair is further hashed, and this process is repeated until only a single hash value, called “root”, remains. This yields a tree with the leaves corresponding to the input blocks and the root corresponding to the last remaining hash value.  $\mathcal{V}$  sends the root to  $\mathcal{P}$ .
- The proving algorithm, **MT.prove**, is run by  $\mathcal{P}$ . It takes a block index,  $i$ , and a tree as inputs. It outputs a vector proof, of  $\log_2(n)$  elements. The proof asserts the membership of  $i$ -th block in the tree, and consists of all the sibling nodes on a path from the  $i$ -th block to the root of the Merkle tree (including  $i$ -th block). The proof is given to  $\mathcal{V}$ .
- The verification algorithm, **MT.verify**, is run by  $\mathcal{V}$ . It takes as an input  $i$ -th block, a proof, and the tree’s root. It checks if the  $i$ -th block corresponds to the root. If the verification passes, it outputs 1; otherwise, it outputs 0.

The Merkle tree-based scheme has two properties: *correctness* and *security*. Informally, the correctness requires that if both parties run the algorithms correctly, then a proof is always accepted by  $\mathcal{V}$ . The security requires that a computationally bounded malicious  $\mathcal{P}$  cannot convince  $\mathcal{V}$  into accepting an incorrect proof, e.g., proof for a non-member block. The security relies on the assumption that it is computationally infeasible to find the hash function’s collision. Usually, for the sake of simplicity, it is assumed that the number of blocks,  $n$ , is a power of 2. The height of the tree, constructed on  $m$  blocks, is  $\log_2(n)$ .

## F Enhanced OLE’s Protocol

The PSIs proposed in [22] use an enhanced version of the OLE. The enhanced OLE ensures that the receiver cannot learn anything about the sender’s inputs, in the case where it sets its input to 0, i.e.,  $c = 0$ . The enhanced OLE’s protocol (denoted by  $\text{OLE}^+$ ) is presented in Figure 5. In this paper,  $\text{OLE}^+$  is used as a subroutine in **VOPR**, in Section 5.1.

- |   |
|---|
| <ol style="list-style-type: none"> <li>1. Receiver (input <math>c \in \mathbb{F}_p</math>): Pick a random value, <math>r \xleftarrow{\\$} \mathbb{F}_p</math>, and send <b>(inputS, <math>(c^{-1}, r)</math>)</b> to the first <math>\mathcal{F}_{\text{OLE}}</math>.</li> <li>2. Sender (input <math>a, b \in \mathbb{F}_p</math>): Pick a random value, <math>u \xleftarrow{\\$} \mathbb{F}_p</math>, and send <b>(inputR, <math>u</math>)</b> to the first <math>\mathcal{F}_{\text{OLE}}</math>, to learn <math>t = c^{-1} \cdot u + r</math>. Send <b>(inputS, <math>(t + a, b - u)</math>)</b> to the second <math>\mathcal{F}_{\text{OLE}}</math>.</li> <li>3. Receiver: Send <b>(inputR, <math>c</math>)</b> to the second <math>\mathcal{F}_{\text{OLE}}</math> and obtain <math>k = (t + a) \cdot c + (b - u) = a \cdot c + b + r \cdot c</math>. Output <math>s = k - r \cdot c = a \cdot c + b</math>.</li> </ol> |
|---|

Fig. 5: Enhanced Oblivious Linear function Evaluation ( $\text{OLE}^+$ ) [22].

## G Coin-Tossing Protocol

A Coin-Tossing protocol,  $\text{CT}$ , allows two mutually distrustful parties, say  $A$  and  $B$ , to jointly generate a single random bit. Formally,  $\text{CT}$  computes the functionality  $f_{\text{CT}}(in_A, in_B) \rightarrow (out_A, out_B)$ , which takes  $in_A$  and  $in_B$  as inputs of  $A$  and  $B$  respectively and outputs  $out_A$  to  $A$  and  $out_B$  to  $B$ , where  $out_A = out_B$ . A basic security requirement of a  $\text{CT}$  is that the resulting bit is (computationally) indistinguishable from a truly random bit.

Blum proposed a simple  $\text{CT}$  in [10] that works as follows. Party  $A$  picks a random bit  $in_A \xleftarrow{\$} \{0, 1\}$ , commits to it and sends the commitment to  $B$  which sends its choice of random input,  $in_B \xleftarrow{\$} \{0, 1\}$ , to  $A$ . Then,  $A$  sends the opening of the commitment (including  $in_A$ ) to  $B$ , which checks whether the commitment matches its opening. If so, each party computes the final random bit as  $in_A \oplus in_B$ .

There have also been *fair* coin-tossing protocols, e.g., in [36], that ensure either both parties learn the result or nobody does. These protocols can be generalised to *multi-party* coin-tossing protocols to generate a *random string* (rather than a single bit), e.g., see [7, 31]. The overall computation and communication complexities of (fair) multi-party coin-tossing protocols are often linear with the number of participants. In this paper, any secure multi-party  $\text{CT}$  that generates a random string can be used. For the sake of simplicity, we let a multi-party  $f_{\text{CT}}$  take  $m$  inputs and output a single value, i.e.,  $f_{\text{CT}}(in_1, \dots, in_m) \rightarrow out$ .

## H Formal Definitions of Predicates of $\mathcal{PSI}^{\mathcal{F}^c}$

In this section, we present the formal definitions of predicates  $Q^{\text{Init}}$ ,  $Q^{\text{Del}}$ ,  $Q^{\text{UF-A}}$  and  $Q^{\text{F-A}}$ . We borrowed the first three predicates from the “fair and robust multi-party computation” initially proposed in [32]; nevertheless, we (i) have introduced an additional predicate  $Q^{\text{F-A}}$  and (ii) will provide more formal accurate definitions of these predicates.

**Definition 7 ( $Q^{\text{Init}}$ : Initiation predicate).** Let  $\mathcal{G}$  be a stable ledger,  $adr_{sc}$  be smart contract  $sc$ ’s address,  $Adr$  be a set of  $m + 1$  distinct addresses, and  $\ddot{x}$  be a fixed amount of coins. Then, predicate  $Q^{\text{Init}}(\mathcal{G}, adr_{sc}, m + 1, Adr, \ddot{x})$  returns 1 if every address in  $Adr$  has at least  $\ddot{x}$  coins in  $sc$ ; otherwise, it returns 0.

**Definition 8 ( $Q^{\text{Del}}$ : Delivery predicate).** Let  $\text{pram} := (\mathcal{G}, adr_{sc}, \ddot{x})$  be the parameters defined above, and  $adr_i \in Adr$  be the address of an honest party. Then, predicate  $Q^{\text{Del}}(\text{pram}, adr_i)$  returns 1 if  $adr_i$  has sent  $\ddot{x}$  amount to  $sc$  and received  $\ddot{x}$  amount from it; thus, its balance in  $sc$  is 0. Otherwise, it returns 0.

**Definition 9 ( $Q^{\text{UF-A}}$ : UnFair-Abort predicate).** Let  $\text{pram} := (\mathcal{G}, adr_{sc}, \ddot{x})$  be the parameters defined above, and  $Adr' \subset Adr$  be a set containing honest parties’ addresses,  $m' = |Adr'|$ , and  $adr_i \in Adr'$ . Let also  $G$  be a compensation function that takes as input three parameters  $(\text{deps}, adr_i, m')$ , where  $\text{deps}$  is the amount of coins that all  $m + 1$  parties deposit. It returns the amount of compensation each honest party must receive, i.e.,  $G(\text{deps}, adr_i, m') \rightarrow \ddot{x}_i$ . Then, predicate



$Q^{UF-A}$  is defined as  $Q^{UF-A}(\text{pram}, G, \ddot{\text{deps}}, m', \text{adr}_i) \rightarrow (a, b)$ , where  $a = 1$  if  $\text{adr}_i$  is an honest party's address and  $\text{adr}_i$  has sent  $\ddot{x}$  amount to  $sc$  and received  $\ddot{x} + \ddot{x}_i$  from it, and  $b = 1$  if  $\text{adr}_i$  is  $Aud$ 's address and  $\text{adr}_i$  received  $\ddot{x}_i$  from  $sc$ . Otherwise,  $a = b = 0$ .

**Definition 10 ( $Q^{F-A}$ : Fair-Abort predicate).** Let  $\text{pram} := (\mathcal{G}, \text{adr}_{sc}, \ddot{x})$  be the parameters defined above, and  $\text{Adr}' \subset \text{Adr}$  be a set containing honest parties' addresses,  $m' = |\text{Adr}'|$ ,  $\text{adr}_i \in \text{Adr}'$ , and  $\text{adr}_j$  be  $Aud$ 's address. Let  $G$  be the compensation function, defined above and let  $G(\text{deps}, \text{adr}_j, m') \rightarrow \ddot{x}_j$  be the compensation that the auditor must receive. Then, predicate  $Q^{F-A}(\text{pram}, G, \text{deps}, m', \text{adr}_i, \text{adr}_j)$  returns 1, if  $\text{adr}_i$  (s.t.  $\text{adr}_i \neq \text{adr}_j$ ) has sent  $\ddot{x}$  amount to  $sc$  and received  $\ddot{x}$  from it, and  $\text{adr}_j$  received  $\ddot{x}_j$  from  $sc$ . Otherwise, it returns 0.

We observed that predicate  $Q^{F-A}$  should have been defined in the generic framework in [32] too; as the framework should have also captured the cases where an adversary may abort without learning any output after the onset of the protocol.

## I Proof of VOPR

*Proof.* Before proving Theorem 1, we present Lemma 1 and Theorem 9 that will be used in the proof of Theorem 1. Informally, Lemma 1 states that the evaluation of a random polynomial at a fixed value results in a uniformly random value.

**Lemma 1.** Let  $x_i$  be an element of a finite field  $\mathbb{F}_p$ , picked uniformly at random and  $\mu(x)$  be a random polynomial of constant degree  $d$  (where  $d = \text{const}(p)$ ) and defined over  $\mathbb{F}_p[X]$ . Then, the evaluation of  $\mu(x)$  at  $x_i$  is distributed uniformly at random over the elements of the field, i.e.,  $\Pr[\mu(x_i) = y] = \frac{1}{p}$ , where  $y \in \mathbb{F}_p$ .

*Proof.* Let  $\mu(x) = a_0 + \sum_{j=1}^d a_j x^j$ , where the coefficients are distributed uniformly at random over the field.

Then, for any choice of  $x$  and random coefficients  $a_1, \dots, a_d$ , it holds that:

$$\Pr[\mu(x) = y] = \Pr\left[\sum_{i=0}^d a_i \cdot x^i = y\right] = \Pr\left[a_0 = y - \sum_{j=1}^d a_j \cdot x^j\right] = \frac{1}{p}$$

$\forall y \in \mathbb{F}_p$ , as  $a_0$  has been picked uniformly at random from  $\mathbb{F}_p$ .

Informally, Theorem 9 states that the product of two arbitrary polynomials (in coefficient form) is a polynomial whose roots are the union of the two original polynomials. Below, we formally state it. The theorem has been taken from [4].

**Theorem 9.** Let  $\mathbf{p}$  and  $\mathbf{q}$  be two arbitrary non-constant polynomials of degree  $d$  and  $d'$  respectively, such that  $\mathbf{p}, \mathbf{q} \in \mathbb{F}_p[X]$  and they are in coefficient form. Then, the product of the two polynomials is a polynomial whose roots include precisely the two polynomials' roots.

We refer readers to Appendix U for the proof of Theorem 9. Next, we prove the main theorem, i.e., Theorem 1, by considering the case where each party is corrupt, in turn.

**Case 1: Corrupt sender.** In the real execution, the sender's view is defined as follows:

$$\text{View}_S^{\text{VOPR}}((\psi, \alpha), \beta) = \{\psi, \alpha, r_s, \beta(z), \theta(z), \text{View}_S^{\text{OLE}^+}, \perp\}$$

where  $r_s$  is the outcome of internal random coins of the sender and  $\text{View}_S^{\text{OLE}^+}$  refers to the sender's real-model view during the execution of  $\text{OLE}^+$ . The simulator  $\text{Sim}_S^{\text{VOPR}}$ , which receives  $\psi$  and  $\alpha$ , works as follows.

1. generates an empty view. It appends to the view polynomials  $(\psi, \alpha)$  and coins  $r'_s$  chosen uniformly at random.
2. computes polynomial  $\beta = \beta_1 \cdot \beta_2$ , where  $\beta_1$  is a random polynomial of degree 1 and  $\beta_2$  is an arbitrary polynomial of degree  $e' - 1$ . Next, it constructs polynomial  $\theta$  as follows:  $\theta = \psi \cdot \beta + \alpha$ .
3. picks value  $z \xleftarrow{\$} \mathbb{F}_p$ . Then, it evaluates polynomials  $\beta$  and  $\theta$  at point  $z$ . This results in values  $\beta_z$  and  $\theta_z$  respectively. It appends these two values to the view.
4. extracts the sender-side simulation of  $\text{OLE}^+$  from  $\text{OLE}^+$ 's simulator. Let  $\text{Sim}_S^{\text{OLE}^+}$  be this simulation. Note, the latter simulation is guaranteed to exist, as  $\text{OLE}^+$  has been proven secure (in [22]). It appends  $\text{Sim}_S^{\text{OLE}^+}$  and  $\perp$  to its view.

Now, we are ready to show that the two views are computationally indistinguishable. The sender's inputs are identical in both models, so they have identical distributions. Since the real-model semi-honest adversary samples its randomness according to the protocol's description, the random coins in both models have identical distributions. Next, we explain why values  $\beta(z)$  in the real model and  $\beta_z$  in the ideal model are (computationally) indistinguishable. In the real model,  $\beta(z)$  is the evaluation of polynomial  $\beta = \beta_1 \cdot \beta_2$  at random point  $z$ , where  $\beta_1$  is a random polynomial. We know that  $\beta(z) = \beta_1(z) \cdot \beta_2(z)$ , for any (non-zero)  $z$ . Moreover, by Lemma 1, we know that  $\beta_1(z)$  is a uniformly random value. Therefore,  $\beta(z) = \beta_1(z) \cdot \beta_2(z)$  is a uniformly random value as well. In the ideal world, polynomial  $\beta$  has the same structure as  $\beta$  has (i.e.,  $\beta = \beta_1 \cdot \beta_2$ , where  $\beta_1$  is a random polynomial). That means  $\beta_z$  is a uniformly random value too. Thus,  $\beta(z)$  and  $\beta_z$  are computationally indistinguishable. Next, we turn our attention to values  $\theta(z)$  in the real model and  $\theta_z$  in the ideal model. We know that  $\theta(z)$  is a function of  $\beta_1(z)$ , as polynomial  $\theta$  has been defined as  $\theta = \psi \cdot (\beta_1 \cdot \beta_2) + \alpha$ . Similarly,  $\theta_z$  is a function of  $\beta_z$ . As we have already discussed,  $\beta(z)$  and  $\beta_z$  are computationally indistinguishable, so are their functions  $\theta(z)$  and  $\theta_z$ . Moreover, as  $\text{OLE}^+$  has been proven secure,  $\text{View}_S^{\text{OLE}^+}$  and  $\text{Sim}_S^{\text{OLE}^+}$  are computationally indistinguishable. It is also clear that  $\perp$  is identical in both models. We conclude that the two views are computationally indistinguishable.

**Case 2: Corrupt receiver.** Let  $\text{Sim}_R^{\text{VOPR}}$  be the simulator, in this case, which uses a subroutine adversary,  $\mathcal{A}_R$ .  $\text{Sim}_R^{\text{VOPR}}$  works as follows.

1. simulates  $\text{OLE}^+$  and receives  $\mathcal{A}_R$ 's input coefficients  $b_j$  for all  $j$ ,  $0 \leq j \leq e'$ , as we are in  $f_{\text{OLE}^+}$ -hybrid model.
2. reconstructs polynomial  $\beta$ , given the above coefficients.
3. simulates the honest sender's inputs as follows. It picks two random polynomials:  $\psi = \sum_{i=0}^e g_i \cdot x^i$  and  $\alpha = \sum_{j=0}^{e+e'} a_j \cdot x^j$ , such that  $g_i \xleftarrow{\$} \mathbb{F}_p$  and every  $a_j$  has the form:  $a_j = \sum_{\substack{k=e' \\ t+k=j}}^{t=e} a_{t,k}$ , where  $t+k=j$  and  $a_{t,k} \xleftarrow{\$} \mathbb{F}_p$ .
4. sends to  $\text{OLE}^+$ 's functionality values  $g_i$  and  $a_{i,j}$  and receives  $c_{i,j}$  from this functionality (for all  $i, j$ ).
5. sends all  $c_{i,j}$  to TTP and receives polynomial  $\theta$ .
6. picks a random value  $z$  from  $\mathbb{F}_p$ . Then, it computes  $\psi_z = \psi(z)$  and  $\alpha_z = \alpha(z)$ .
7. sends  $z$  and all  $c_{i,j}$  to  $\mathcal{A}_R$  which sends back  $\theta_z$  and  $\beta_z$  to the simulator.
8. sends  $\psi_z$  and  $\alpha_z$  to  $\mathcal{A}_R$ .
9. checks if the following relation hold:
$$\beta_z = \beta(z) \quad \wedge \quad \theta_z = \theta(z) \quad \wedge \quad \theta(z) = \psi_z \cdot \beta_z + \alpha_z \quad (2)$$
 If Relation 2 does not hold, it aborts (i.e., sends abort signal  $A$  to the sender) and still proceeds to the next step.
10. outputs whatever  $\mathcal{A}_R$  outputs.

We first focus on the adversary's output. Both values of  $z$  in the real and ideal models have been picked uniformly at random from  $\mathbb{F}_p$ ; therefore, they have identical distributions. In the real model, values  $\psi_z$  and  $\alpha_z$  are the result of the evaluations of two random polynomials at (random) point  $z$ . In the ideal model, values  $\psi_z$  and  $\alpha_z$  are also the result of the evaluations of two random polynomials (i.e.,  $\psi$  and  $\alpha$ ) at point  $z$ . By Lemma 1, we know that the evaluation of a random polynomial at an arbitrary value yields a uniformly random value in  $\mathbb{F}_p$ . Therefore, the distribution of pair  $(\psi_z, \alpha_z)$  in the real model is identical to that of pair  $(\psi_z, \alpha_z)$  in the ideal model. Moreover, the final result (i.e., values  $c_{i,j}$ ) in the real model has the same distribution as the final result (i.e., values  $c_{i,j}$ ) in the ideal model, as they are the outputs of the ideal calls to  $f_{\text{OLE}^+}$ , as we are in the  $f_{\text{OLE}^+}$ -hybrid model.

Next, we turn our attention to the sender's output. We will show that the output distributions of the honest sender in the ideal and real models are statistically close. Our focus will be on the probability that it aborts in each model, as it does not receive any other output. In the ideal model,  $\text{Sim}_R^{\text{VOPR}}$  is given the honestly generated result polynomial  $\theta$  (computed by TTP) and the adversary's input polynomial  $\beta$ .  $\text{Sim}_R^{\text{VOPR}}$  aborts with a probability of 1 if Relation 2 does not hold. However, in the real model, the honest sender (in addition to its inputs) is given only  $\beta_z$  and  $\theta_z$  and is not given polynomials  $\beta$  and  $\theta$ ; it wants to check if the following equation holds,  $\theta_z = \psi_z \cdot \beta_z + \alpha_z$ . Note, polynomial  $\theta = \psi \cdot \beta + \alpha$  (resulted from  $c_{i,j}$ ) is well-structured, as it satisfies the following three conditions, regardless of the adversary's input  $\beta$  to  $\text{OLE}^+$ , (i)

$\deg(\theta) = \text{Max}(\deg(\beta) + \deg(\psi), \deg(\alpha))$ , as  $\mathbb{F}_p[X]$  is an integral domain and  $(\psi, \alpha)$  are random polynomials, (ii) the roots of the product polynomial  $\nu = \psi \cdot \beta$  contains exactly both polynomials' roots, by Theorem 9, and (iii) the roots of  $\nu + \alpha$  is the intersection of the roots of  $\nu$  and  $\alpha$ , as shown in [33]. Furthermore, polynomial  $\theta$  reveals no information (about  $\psi$  and  $\alpha$  except their degrees) to the adversary and the pair  $(\psi_z, \alpha_z)$  is given to the adversary after it sends the pair  $(\theta_z, \beta_z)$  to the sender. There are exactly four cases where pair  $(\theta_z, \beta_z)$  can be constructed by the real-model adversary. Below, we state each case and the probability that the adversary is detected in that case during the verification, i.e.,  $\theta_z \stackrel{?}{=} \psi_z \cdot \beta_z + \alpha_z$ .

1.  $\theta_z = \theta(z) \wedge \beta_z = \beta(z)$ . This is a trivial non-interesting case, as the adversary has behaved honestly, so it can always pass the verification.
2.  $\theta_z \neq \theta(z) \wedge \beta_z = \beta(z)$ . In this case, the adversary is detected with a probability of 1.
3.  $\theta_z = \theta(z) \wedge \beta_z \neq \beta(z)$ . In this case, the adversary is also detected with a probability of 1.
4.  $\theta_z \neq \theta(z) \wedge \beta_z \neq \beta(z)$ . In this case, the adversary is detected with an overwhelming probability, i.e.,  $1 - \frac{1}{2^{2\lambda}}$ .

As we illustrated above, in the real model, the lowest probability that the honest sender would abort in case of adversarial behaviour is  $1 - \frac{1}{2^{2\lambda}}$ . Thus, the honest sender's output distributions in the ideal and real models are statistically close, i.e., 1 vs  $1 - \frac{1}{2^{2\lambda}}$ .

We conclude that the distribution of the joint outputs of the honest sender and adversary in the real and ideal models are computationally indistinguishable.

## J Proof of ZSPA

*Proof.* For the sake of simplicity, we will assume the sender, which generates the result, sends the result directly to the rest of the parties, i.e., receivers, instead of sending it to a smart contract. We first consider the case in which the sender is corrupt.

**Case 1: Corrupt sender.** Let  $\text{Sim}_S^{\text{ZSPA}}$  be the simulator using a subroutine adversary,  $\mathcal{A}_S$ .  $\text{Sim}_S^{\text{ZSPA}}$  works as follows.

1. simulates CT and receives the output value  $k$  from  $f_{\text{CT}}$ , as we are in  $f_{\text{CT}}$ -hybrid model.
2. sends  $k$  to TTP and receives back from it  $m$  pairs, where each pair is of the form  $(g, q)$ .
3. sends  $k$  to  $\mathcal{A}_S$  and receives back from it  $m$  pairs where each pair is of the form  $(g', q')$ .
4. checks whether the following equations hold (for each pair):  $g = g' \wedge q = q'$ . If the two equations do not hold, then it aborts (i.e., sends abort signal  $\perp$  to the receiver) and proceeds to the next step.
5. outputs whatever  $\mathcal{A}_S$  outputs.

We first focus on the adversary's output. In the real model, the only messages that the adversary receives are those messages it receives as the result of the ideal call to  $f_{\text{ct}}$ . These messages have identical distribution to the distribution of the messages in the ideal model, as the CT is secure. Now, we move on to the receiver's output. We will show that the output distributions of the honest receiver in the ideal and real models are computationally indistinguishable. In the real model, each element of pair  $(g, p)$  is the output of a deterministic function on the output of  $f_{\text{ct}}$ . We know the output of  $f_{\text{ct}}$  in the real and ideal models have an identical distribution, and so do the evaluations of deterministic functions (i.e., Merkle tree, H, and PRF) on them, as long as these three functions' correctness hold. Therefore, each pair  $(g, q)$  in the real model has an identical distribution to pair  $(g, q)$  in the ideal model. For the same reasons, the honest receiver in the real model aborts with the same probability as  $\text{Sim}_S^{\text{ZSPA}}$  does in the ideal model. We conclude that the distributions of the joint outputs of the adversary and honest receiver in the real and ideal models are (computationally) indistinguishable.

**Case 2: Corrupt receiver.** Let  $\text{Sim}_R^{\text{ZSPA}}$  be the simulator that uses subroutine adversary  $\mathcal{A}_R$ .  $\text{Sim}_R^{\text{ZSPA}}$  works as follows.

1. simulates CT and receives the output value  $k$  from  $f_{\text{ct}}$ .
2. sends  $k$  to TTP and receives back  $m$  pairs of the form  $(g, q)$  from TTP.
3. sends  $(k, g, q)$  to  $\mathcal{A}_R$  and outputs whatever  $\mathcal{A}_R$  outputs.

In the real model, the adversary receives two sets of messages, the first set includes the transcripts (including  $k$ ) it receives when it makes an ideal call to  $f_{\text{ct}}$  and the second set includes pair  $(g, q)$ . As we already discussed above (because we are in the  $f_{\text{ct}}$ -hybrid model) the distributions of the messages it receives from  $f_{\text{ct}}$  in the real and ideal models are identical. Moreover, the distribution of  $f_{\text{ct}}$ 's output (i.e.,  $\bar{k}$  and  $k$ ) in both models is identical; therefore, the honest sender's output distribution in both models is identical. As we already discussed, the evaluations of deterministic functions (i.e., Merkle tree, H, and PRF) on  $f_{\text{ct}}$ 's outputs have an identical distribution. Therefore, each pair  $(g, q)$  in the real model has an identical distribution to the pair  $(g, q)$  in the ideal model. Hence, the distribution of the joint outputs of the adversary and honest receiver in the real and ideal models is indistinguishable.

In addition to the security guarantee (i.e., computation's correctness against malicious sender or receiver) stated by Theorem 2, ZSPA offers (a) privacy against the public, and (b) non-refutability. Informally, privacy here means that given the state of the contract (i.e.,  $g$  and  $q$ ), an external party cannot learn any information about any of the pseudorandom values,  $z_j$ ; while non-refutability means that if a party sends "approved" then in future cannot deny the knowledge of the values whose representation is stored in the contract.

**Theorem 10.** *If H is preimage resistance, PRF is secure, the signature scheme used in the smart contract is secure (i.e., existentially unforgeable under chosen message attacks), and the blockchain is secure (i.e., offers persistence and*

*liveness properties [20]) then ZSPA offers (i) privacy against the public and (ii) non-refutability.*

*Proof.* First, we focus on privacy. Since key  $k$ , for PRF, has been picked uniformly at random and  $H$  is preimage resistance, the probability that given  $g$  the adversary can find  $k$  is negligible in the security parameter, i.e.,  $\epsilon(\lambda)$ . Furthermore, because PRF is secure (i.e., its outputs are indistinguishable from random values) and  $H$  is preimage resistance, given the Merkle tree's root  $g$ , the probability that the adversary can find a leaf node, which is the output of PRF, is  $\epsilon(\lambda)$  too.

Now we move on the non-refutability. Due to the persistency property of the blockchain, once a transaction/message goes more than  $v$  blocks deep into the blockchain of one honest player (where  $v$  is a security parameter), then it will be included in every honest player's blockchain with overwhelming probability, and it will be assigned a permanent position in the blockchain (so it will not be modified with an overwhelming probability). Also, due to the liveness property, all transactions originating from honest parties will eventually end up at a depth of more than  $v$  blocks in an honest player's blockchain; therefore, the adversary cannot perform a selective denial of service attack against honest account holders. Moreover, due to the security of the digital signature (i.e., existentially unforgeable under chosen message attacks), one cannot deny sending the messages it sent to the blockchain and smart contract.

## K Proof of ZSPA-A

*Proof.* First, we consider the case where a sender, who (may collude with  $m - 2$  senders and) generates pairs  $(g, q)$ , is corrupt.

**Case 1: Corrupt sender.** Let  $\text{Sim}_s^{\text{ZSPA-A}}$  be the simulator using a subroutine adversary,  $\mathcal{A}_s$ . Below, we explain how  $\text{Sim}_s^{\text{ZSPA-A}}$  works.

1. simulates CT and receives the output value  $k$  from  $f_{\text{CT}}$ .
2. sends  $k$  to TTP and receives back from it  $m$  pairs, where each pair is of the form  $(g, q)$ .
3. sends  $k$  to  $\mathcal{A}_s$  and receives back from it  $m$  pairs where each pair is of the form  $(g', q')$ .
4. constructs an empty vector  $L$ .  $\text{Sim}_s^{\text{ZSPA-A}}$  checks whether the following equations hold for each  $j$ -th pair:  $g = g' \wedge q = q'$ . If these two equations do not hold, it sends an abort message  $A$  to other receiver clients, appends the index of the pair (i.e.,  $j$ ) to  $L$ , and proceeds to the next step for the valid pairs. In the case where there are no valid pairs, it moves on to step 9.
5. picks a random polynomial  $\zeta$  of degree 1. Moreover, for every  $j \notin L$ ,  $\text{Sim}_s^{\text{ZSPA-A}}$  picks a random polynomial  $\xi^{(j)}$  of degree  $b' - 1$ , where  $1 \leq j \leq m$ .
6. computes  $m$  pseudorandom values for every  $i, j'$ , where  $0 \leq i \leq b'$  and  $j' \notin L$  as follows.

$$\forall j', 1 \leq j' \leq m - 1 : z_{i,j} = \text{PRF}(k, i || j') \quad \text{and} \quad z_{i,m} = - \sum_{j'=1}^{m-1} z_{i,j}$$

7. generates polynomial  $\mu^{(j)}$ , for every  $j \notin L$ , as follows:  $\mu^{(j)} = \zeta \cdot \xi^{(j)} - \tau^{(j)}$ ,  
where  $\tau^{(j)} = \sum_{i=0}^{b'} z_{i,j} \cdot x^i$ .
8. sends the above  $\zeta$ ,  $\xi^{(j)}$ , and  $\mu^{(j)}$  to all parties (i.e.,  $\mathcal{A}_S$  and the receivers),  
for every  $j \notin L$ .
9. outputs whatever  $\mathcal{A}_S$  outputs.

Now, we focus on the adversary's output. In the real model, the messages that the adversary receives include those messages it receives as the result of the ideal call to  $f_{\text{ct}}$  and  $(\zeta, \xi^{(j)}, \mu^{(j)})$ , where  $j \notin L$  and  $1 \leq j \leq m$ . Those messages yielded from the ideal calls have identical distribution to the distribution of the messages in the ideal model, as  $\text{CT}$  is secure. The distribution of each  $\mu^{(j)}$  depends on the distribution of its components; namely,  $\zeta$ ,  $\xi^{(j)}$ , and  $\tau^{(j)}$ . As we are in the  $f_{\text{ct}}$ -hybrid model, the distributions of  $\tau^{(j)}$  in the real model and  $\tau^{(j)}$  in the ideal model are identical, as they were derived from the output of  $f_{\text{ct}}$ . Furthermore, in the real model, each polynomial  $\zeta$  and  $\xi^{(j)}$  has been picked uniformly at random and they are independent of the clients' and the adversary's inputs. The same arguments hold for  $(\zeta, \xi^{(j)}, \mu^{(j)})$  in the ideal model. Therefore,  $(\zeta, \xi^{(j)}, \mu^{(j)})$  in the real model and  $(\zeta, \xi^{(j)}, \mu^{(j)})$  in the ideal model have identical distributions.

Next, we turn our attention to the receiver's output. We will show that the output distributions of an honest receiver and the auditor in the ideal and real models are computationally indistinguishable. In the real model, each element of the pair  $(g, p)$  is the output of a deterministic function on the output of  $f_{\text{ct}}$ . We know the outputs of  $f_{\text{ct}}$  in the real and ideal models have an identical distribution, and so do the evaluations of deterministic functions (namely Merkle tree,  $\text{H}$ , and  $\text{PRF}$ ) on them. Therefore, each pair  $(g, q)$  in the real model has an identical distribution to the pair  $(g, q)$  in the ideal model. For the same reasons, the honest receiver in the real model aborts with the same probability as  $\text{Sim}_S^{\text{ZSPA-A}}$  does in the ideal model. The same argument holds for the arbiter's output, as it performs the same checks that an honest receiver does. Thus, the distribution of the joint outputs of the adversary, honest receiver, and honest in the real and ideal models is computationally indistinguishable.

**Case 2: Corrupt receiver.** Let  $\text{Sim}_R^{\text{ZSPA-A}}$  be the simulator that uses subroutine adversary  $\mathcal{A}_R$ . Below, we explain how  $\text{Sim}_R^{\text{ZSPA-A}}$  works.

1. simulates  $\text{ZSPA}$  and receives the  $m$  output pairs of the form  $(k, g, q)$  from  $f^{\text{ZSPA}}$ .
2. sends  $(k, g, q)$  to  $\mathcal{A}_R$  and receives  $m$  keys,  $k'_j$ , where  $1 \leq j \leq m$ .
3. generates an empty vector  $L$ . Next, for every  $j$ ,  $\text{Sim}_R^{\text{ZSPA-A}}$  computes  $q'_j$  as  $\text{H}(k'_j) = q_j$ . It generates  $g_j$  as follows.
  - (a) for every  $i$  (where  $0 \leq i \leq b'$ ), generates  $m$  pseudorandom values as below.

$$\forall j, 1 \leq j' \leq m-1 : z_{i,j} = \text{PRF}(k'_j, i || j'), \quad z_{i,m} = - \sum_{j=1}^{m-1} z_{i,j}$$

- (b) constructs a Merkle tree on top of all pseudorandom values,  
 $\text{MT.genTree}(z_{0,1}, \dots, z_{b',m}) \rightarrow g'_j$ .
4. checks if the following equations hold for each  $j$ -th pair:  $(k = k'_j) \wedge (g = g'_j) \wedge (q = q'_j)$ .
  5. If these equations do not hold for  $j$ -th value, it appends  $j$  to  $L$  and proceeds to the next step for the valid value. In the case where there is no valid value, it moves on to step 9.
  6. picks a random polynomial  $\zeta$  of degree 1. Also, for every  $j \notin L$ , it picks a random polynomial  $\xi^{(j)}$  of degree  $b' - 1$ , where  $1 \leq j \leq m$ .
  7. generates polynomial  $\mu^{(j)}$ , for every  $j \notin L$ , as follows:  $\mu^{(j)} = \zeta \cdot \xi^{(j)} - \tau^{(j)}$ ,  
where  $\tau^{(j)} = \sum_{i=0}^{b'} z_{i,j} \cdot x^i$ , and values  $z_{i,j}$  were generated in step 3a.
  8. sends the above  $\zeta$ ,  $\xi^{(j)}$ , and  $\mu^{(j)}$  to  $\mathcal{A}_R$ , for every  $j \notin L$  and  $1 \leq j \leq m$ .
  9. outputs whatever  $\mathcal{A}_R$  outputs.

In the real model, the adversary receives two sets of messages, the first set includes the transcripts (including  $k, g, q$ ) it receives when it makes an ideal call to  $f^{\text{ZSPA}}$  and the second set includes pairs  $(\zeta, \xi^{(j)}, \mu^{(j)})$ , for every  $j \notin L$  and  $1 \leq j \leq m$ . Since we are in the  $f^{\text{ZSPA}}$ -hybrid model and (based on our assumption) there is at least one honest party participated in ZSPA (i.e., there are at most  $m - 1$  malicious participants of ZSPA), the distribution of the messages it receives from  $f^{\text{ZSPA}}$  in the real and ideal models is identical. Furthermore, as we discussed in Case 1,  $(\zeta, \xi^{(j)}, \mu^{(j)})$  in the real model and  $(\zeta, \xi^{(j)}, \mu^{(j)})$  in the ideal model have identical distribution. The honest sender's output distribution in both models is identical, as the distribution of  $f_{\text{CT}}$ 's output (i.e.,  $k$ ) in both models is identical.

Now we show that the probability that the auditor aborts in the ideal and real models are statistically close. In the ideal model,  $\text{Sim}_R^{\text{ZSPA-A}}$  is given the ideal functionality's output that includes key  $k$ . Therefore, it can check whether the key that  $\mathcal{A}_R$  sends to it equals  $k$ , i.e.,  $k \stackrel{?}{=} k'_j$ . Thus, it aborts with the probability 1. However, in the real model, an honest auditor is not given the output of CT (say key  $k$ ) and it can only check whether the key is consistent with the hash value  $q$  and the Merkle tree's root  $g$  stored on the blockchain. This means the adversary can distinguish the two models if in the real model it sends a key  $\tilde{k}$ , such that  $\tilde{k} \neq k$  and still passes the checks. Specifically, it sends the invalid key  $\tilde{k}$  that can generate valid pair  $(g, q)$ , as follows:  $\text{H}(\tilde{k}) = q$  and  $\text{MT.genTree}(z'_{0,1}, \dots, z'_{b',m}) \rightarrow g$ , where each  $z'_{i,j}$  is derived from  $\tilde{k}$  using the same technique described in step 3 above. Nevertheless, this means that the adversary breaks the second preimage resistance property of  $\text{H}$ ; however,  $\text{H}$  is the second-preimage resistance and the probability that the adversary succeeds in finding the second preimage is negligible in the security parameter, i.e.,  $\epsilon(\lambda)$  where  $\lambda$  is the hash function's security parameter. Therefore, in the real model, the auditor aborts if an invalid key is provided with a probability  $1 - \epsilon(\lambda)$  which is statically close to the probability that  $\text{Sim}_R^{\text{ZSPA-A}}$  aborts in the same situation in the ideal model, i.e.,  $1 - \epsilon(\lambda)$  vs 1. Hence, the distribution of the joint outputs of the adversary, honest sender, and honest auditor in the real and ideal models is indistinguishable.



## L Proof of Unforgeable Polynomials

We first prove Theorem 4, i.e., unforgeable polynomials.

*Proof.* Let  $\tau = \delta - \gamma$  and  $\zeta = a \cdot x + b$ . Since  $\gamma$  is a random polynomial of degree  $d + 1$  and unknown to the adversary, given  $(\theta, \pi)$ , the adversary cannot learn anything about the factor  $\zeta$ ; as from its point of view every polynomial of degree 1 in  $\mathbb{F}_p[X]$  is equally likely to be  $\zeta$ . Moreover, polynomial  $\tau$  has at most  $\text{Max}(\text{deg}(\delta), d + 1)$  irreducible non-constant factors. For  $\zeta$  to divide  $\tau$ , one of the factors of  $\tau$  must be equal to  $\zeta$ . We also know that  $\zeta$  has been picked uniformly at random (i.e.,  $a, b \xleftarrow{\$} \mathbb{F}_p$ ). Thus, the probability that  $\zeta$  divides  $\tau$  is negligible in the security parameter,  $\lambda$ . Specifically,

$$\Pr[\zeta \mid (\delta - \gamma)] \leq \frac{\text{Max}(\text{deg}(\delta), d + 1)}{2^\lambda} = \epsilon(\lambda)$$

Next, we prove Theorem 5, i.e., unforgeable polynomials' linear combination.

*Proof.* This proof is a generalisation of that of Theorem 4. Let  $\tau_j = \delta_j - \gamma_j$  and  $\zeta = a \cdot x + b$ . Since every  $\gamma_j$  is a random polynomial of degree  $d + 1$  and unknown to the adversary, given  $(\theta, \vec{\pi})$ , the adversary cannot learn anything about the factor  $\zeta$ . Each polynomial  $\tau_j$  has at most  $\text{Max}(\text{deg}(\delta_j), d + 1)$  irreducible non-constant factors. We know that  $\zeta$  has been picked uniformly at random (i.e.,  $a, b \xleftarrow{\$} \mathbb{F}_p$ ). Therefore, the probability that  $\zeta$  divides  $\sum_{j=1}^t \delta_j + \sum_{j=t+1}^n \theta_j - \sum_{j=1}^n \gamma_j$  equals the probability that  $\zeta$  equals to one of the factors of every  $\tau_j$ , that is negligible in the security parameter. Concretely,

$$\Pr[\zeta \mid (\sum_{j=1}^t \delta_j + \sum_{j=t+1}^n \theta_j - \sum_{j=1}^n \gamma_j)] \leq \frac{\prod_{j=1}^t \text{Max}(\text{deg}(\delta_j), d + 1)}{2^{\lambda t}} = \epsilon(\lambda)$$

## M Workflow of Justitia

Figure 6 outlines the interaction between parties in Justitia.

## N Further Discussion on Justitia

### N.1 Input and Output Privacy

Intuitively, all parties locally blind/encrypt all (set element related) messages before they send them the smart contract  $\mathcal{SC}_{\text{JUS}}$ . Also, the resulting polynomial that  $\mathcal{SC}_{\text{JUS}}$  computes is in a blinded form. Thus, (i) no party (including the protocol's participants) can learn other parties input set elements and (ii) the non-participants of the protocol cannot learn anything about the protocol's output, not even the intersection cardinality.

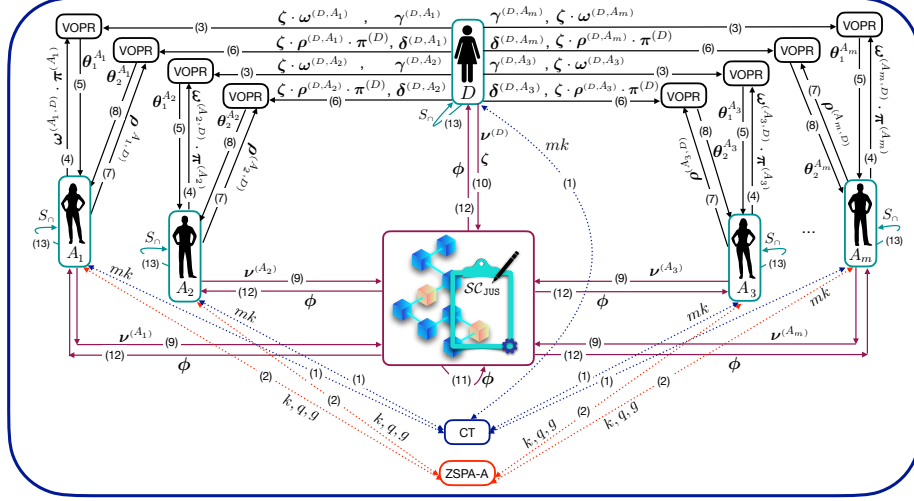


Fig. 6: Outline of the interactions between parties in Justitia

## N.2 Strawman Approaches

**Relying on a Server-aided PSI.** One may be tempted to replace *Justitia* with a scheme in which all clients send their encrypted sets to a server (potentially semi-honest and plays *Aud*'s role) which computes the result in a privacy-preserving manner. We highlight that the main difference is that in this (hypothetical) scheme the server is *always involved*; whereas, in our protocol, *Aud* remains offline as long as the clients behave honestly and it is invoked only when the contract detects misbehaviours.

**Charging the Buyer a Flat Fee.** One may want to add a straightforward payment mechanism to an existing multi-party PSI in such a way that a buyer always must pay a fixed amount, e.g., depending on the total number of clients and the minimum size of the sets. However, this approach is problematic, because it:

1. forces the buyer to always pay even if some malicious clients misbehave during the protocol execution and affect the results' correctness which would allow malicious clients to learn the result at the buyer's expense without letting it learn the correct result, as there exists no fair multi-party PSI in the literature to ensure either all parties learn a correct result or neither does.
2. forces the buyer to always pay independent of the exact size of the intersection. If it has to pay more than the amount it would have paid for the exact cardinality of the intersection then the buyer would be discouraged to

participate in the protocol in the first place. On the other hand, if the buyer has to pay less than what it would have paid for the size of the intersection, then other clients would be discouraged to participate in the protocol.

### N.3 Error Probability

Recall that in JUS, in step 5, each client  $C$  needs to ensure polynomials  $\omega^{(C,D)}$ ,  $\pi^{(C)}$  and  $\rho^{(C,D)}$  do not contain any zero coefficient. This check ensures that client  $C$  (the receiver in VOPR) does not insert any zero values to VOPR (in particular to  $\text{OLE}^+$  which is a subroutine of VOPR). If a zero value is inserted in VOPR, then an honest receiver will learn only a random value and more importantly cannot pass VOPR's verification phase. Nevertheless, this check can be removed from step 5, if we allow JUS to output an error with a small probability.<sup>4</sup> In the remainder of this section, we show that this probability is negligible.

First, we focus on the product  $\omega^{(C,D)} \cdot \pi^{(C)}$ . We know that  $\pi^{(C)}$  is of the form  $\prod_{i=1}^d (x - s'_i)$ , where  $s'_i$  is either a set element  $s_i$  or a random value. Thus, all of its coefficients are non-zero. Also, the probability that at least one of the coefficients of  $d$ -degree random polynomial  $\omega^{(C,D)}$  equals 0 is at most  $\frac{d+1}{p}$ . Below, we state it formally.

**Theorem 11.** *Let  $\delta = \sum_{t=0}^d u_t \cdot x^t$ , where  $u_t \xleftarrow{\$} \mathbb{F}_p$ , for all  $t, 0 \leq t \leq d$ . Then, the probability that at least one of the coefficients equals 0 is at most  $\frac{d+1}{p}$ , i.e.,*

$$\Pr[\exists u_t, u_t = 0] \leq \frac{d+1}{p}$$

*Proof.* The proof is straightforward. Since  $\delta$ 's coefficients are picked uniformly at random from  $\mathbb{F}_p$ , the probability that  $u_t$  equals 0 is  $\frac{1}{p}$ . Since  $\delta$  is of degree  $d$ , due to the union bound, the probability that at least one of  $u_t$ s equals 0 is at most  $\frac{d+1}{p}$ .

Next, we show that the probability that the polynomial  $\omega^{(C,D)} \cdot \pi^{(C)}$  has at least one zero coefficient is negligible in the security parameter; we assume polynomial  $\omega^{(C,D)}$  has no zero coefficient.

**Theorem 12.** *Let  $\alpha = \sum_{i=0}^m a_i \cdot x^i$  and  $\beta = \sum_{j=0}^n b_j \cdot x^j$ , where  $a_i \xleftarrow{\$} \mathbb{F}_p$  and  $a_i, b_j \neq 0$ , for all  $i, j, 0 \leq i \leq m$  and  $0 \leq j \leq n$ . Also, let  $\gamma = \alpha \cdot \beta = \sum_{j=0}^{m+n} c_j \cdot x^j$ . Then, the probability that at least one of the coefficients of polynomial  $\gamma$  equals 0 is at most  $\frac{m+n+1}{p}$ , i.e.,*

$$\Pr[\exists c_j, c_j = 0] \leq \frac{m+n+1}{p}$$

---

<sup>4</sup> By error we mean even if all parties are honest, VOPR halts when an honest party inserts 0 to it.

*Proof.* Each coefficient  $c_k$  of  $\gamma$  can be defined as  $c_k = \sum_{\substack{j=0 \\ i=0}}^{i=m, j=n} a_i \cdot b_j$ , where  $i + j = k$ .

We can rewrite  $c_k$  as  $c_k = a_w \cdot b_z + \sum_{\substack{j=0, j \neq z \\ i=0, i \neq w}}^{i=m, j=n} a_i \cdot b_j$ , where  $w + z = i + j = k$ . We consider two cases for each  $c_k$ :

- Case 1:  $\sum_{\substack{j=0, j \neq z \\ i=0, i \neq w}}^{i=m, j=n} a_i \cdot b_j = 0$ . This is a trivial case, because with the probability of 1 it holds that  $c_k = a_w \cdot b_z \neq 0$ , as by definition  $a_w, b_z \neq 0$  and  $\mathbb{F}_p$  is an integral domain.
- Case 2:  $q = \sum_{\substack{j=0, j \neq z \\ i=0, i \neq w}}^{i=m, j=n} a_i \cdot b_j \neq 0$ . In this case, for event  $c_k = a_w \cdot b_z + q = 0$  to occur,  $a_w \cdot b_z$  must equal the additive inverse of  $q$ . Since  $a_w$  has been picked uniformly at random, the probability that such an even occurs is  $\frac{1}{p}$ .

The above analysis is for a single  $c_k$ . Thus, due to the union bound, the probability that at least one of the coefficients  $c_k$  equals 0 is at most  $\sum_{j=0}^{m+n} \frac{1}{p} = \frac{m+n+1}{p}$ .

Next we turn our attention to  $\rho^{(C,D)}$ . Due to Theorem 11, the probability that at least one of the coefficients of  $\rho^{(C,D)}$  equals 0 is at most  $\frac{d+1}{p}$ .

Hence, due to Theorems 11, 12, and union bound, the probability that at least one of the coefficients in  $\omega^{(C,D)} \cdot \pi^{(C)}$  and  $\rho^{(C,D)}$  equals 0 is at most  $\frac{3d+2}{p}$ , which is negligible in  $p$ .

#### N.4 Main Challenges that Justitia Overcomes

We needed to address several key challenges, to design an efficient scheme that realises  $\mathcal{PSI}^{\mathcal{FC}}$ . Below, we outline these challenges.

**Keeping Overall Complexities Low.** In general, in multi-party PSIs, each client needs to send messages to the rest of the clients and/or engage in secure computation with them, e.g., in [26,35], which would result in communication and/or computation quadratic with the number of clients. To address this challenge, we (a) allow one of the clients as a dealer to interact with the rest of the clients,<sup>5</sup> and (b) we use a smart contract, which acts as a bulletin board to which most messages are sent and also performs lightweight computation on the clients' messages. The combination of these approaches will keep the overall communication and computation linear with the number of clients (and sets' cardinality).

<sup>5</sup> This approach has similarity with the non-secure PSIs in [22].

**Randomising Input Polynomials.** In multi-party PSIs that use the polynomial representation, it is essential that a client’s input polynomial be randomised by another client [4]. To do that securely and efficiently, we require the dealer and each client together to engage in an instance of VOPR, which we developed in Section 5.

**Preserving the Privacy of Outgoing Messages.** Although the use of public smart contracts (e.g., Ethereum) will help keep overall complexity low, it introduces another challenge; namely, if clients do not protect the privacy of the messages they send to the smart contracts, then other clients (e.g., dealer) and non-participants of PSI (i.e., the public) can learn the clients’ set elements and/or the intersection. To efficiently protect the privacy of each client’s messages (sent to the contracts) from the dealer, we require the clients (except the dealer) to engage in ZSPA-A which lets each of them generate a pseudorandom polynomial with which it can blind its message. To protect the privacy of the intersection from the public, we require all clients to run a coin-tossing protocol to agree on a blinding polynomial, with which the final result that encodes the intersection on the smart contract will be blinded.

**Ensuring the Correctness of Subroutine Protocols’ Outputs.** In general, any MPC that must remain secure against an active adversary is equipped with a verification mechanism that ensures an adversary is detected if it affects messages’ integrity, during the protocol’s execution. This is the case for the subroutine protocols that we use, i.e., VOPR and ZSPA-A. However, this type of check itself is not always sufficient. Because in certain cases, the output of an MPC may be fed as input to another MPC and we need to ensure that the *intact* output of the first MPC is fed to the second one. This is the case in our PSI’s subroutines too. To address this challenge, we use unforgeable polynomials; specifically, the output of VOPR is an unforgeable polynomial (that encodes the actual output). If the adversary tampers with the VOPR’s output and uses it later, then a verifier can detect it. We will have the same integrity guarantee for the output of ZSPA-A for free. Because (i) VOPR is called before ZSPA-A, and (ii) if clients use intact outputs of ZSPA-A, then the final result (i.e., the sum of all clients’ messages) will not contain any output of ZSPA-A, as they will cancel out each other. Thus, by checking the correctness of the final result, one can ensure the correctness of the outputs of VOPR and ZSPA-A, in one go.

**Confidentiality of polynomial  $\zeta$ .** The dealer is the only party who picks and knows the secret polynomial  $\zeta$ . This polynomial remains secret until all parties provide their messages to the smart contract. Then, the dealer in step 10 on page 18 sends  $\zeta$  to the smart contract. After this point it is not secret, however, the know of  $\zeta$  would not benefit malicious parties at this point, as they have all provided their inputs to the smart contract.

## N.5 Concrete Parameters

**Hash Table Parameters.** As stated in Appendix D, a hash table has the following main parameters: (1)  $h$ : the number of bins, (2)  $d$ : the bin's maximum size (or capacity), (3)  $c$ : the maximum number of elements that are mapped to the hash table, and (4)  $pr$ : the probability that the number of elements mapped to a bin does not exceed a predefined capacity. Note that in the context of PSI,  $c$  is the maximum of the sizes of all sets.

The literature (e.g., in [3,35,41]) has already studied the concrete parameters of a hash table, even in the context of PSI. For instance, as demonstrated in Section 6 and Appendix J.1 in [3], when  $c$  is in the range  $[2^{10}, 2^{20}]$  and  $pr = 2^{-40}$ , then  $d = 100$ . Furthermore, we have  $h \approx \frac{4c}{d}$ . To provide a concrete value, we would have  $h = 41943$ , when  $c = 2^{20}$  and  $d = 100$ .

**Field Size.** In this paper, all arithmetic operations are defined over a finite field  $\mathbb{F}_p$ , where  $\log_2(p) = \lambda$  represents the security parameter. The outputs of  $\text{PRF}(\cdot)$  and  $\text{PRP}(\cdot)$  are also of size  $\lambda$ . Concrete value of  $\lambda$  can be set based on the maximum bit size of set elements. For instance, one can choose  $\lambda = 60$  and  $\lambda = 128$ , when the maximum bit size of set elements is slightly less than 60 and 128 respectively.

## O Proof of JUS

In this section, we prove Theorem 6, i.e., the security of JUS.

*Proof.* We prove Theorem 6 by considering the case where each party is corrupt, at a time.

**Case 1: Corrupt  $m - 1$  clients in  $\{A_1, \dots, A_m\}$ .** Let  $P'$  be a set of at most  $m - 1$  corrupt clients, where  $P' \subset \{A_1, \dots, A_m\}$ . Let set  $\hat{P}$  be defined as follows:  $\hat{P} = \{A_1, \dots, A_m\} \setminus P'$ . Also, let  $\text{Sim}_A^{\text{JUS}}$  be the simulator, which uses a subroutine adversary,  $\mathcal{A}$ . Below, we explain how  $\text{Sim}_A^{\text{JUS}}$  (which receives the input sets of honest dealer  $D$  and honest client(s) in  $\hat{P}$ ) works.

1. constructs and deploys a smart contract. It sends the contract's address to  $\mathcal{A}$ .
2. simulates CT and receives the output value,  $mk$ , from its functionality,  $f_{\text{ct}}$ .
3. simulates ZSPA-A for each bin and receives the output value,  $(k, g, q)$ , from its functionality,  $f^{\text{ZSPA-A}}$ .
4. deposits in the contract the total amount of  $(\ddot{y} + \ddot{c}\ddot{h}) \cdot (m - |P'| + 1)$  coins on behalf of  $D$  and honest client(s) in  $\hat{P}$ . It sends to  $\mathcal{A}$  the amount deposited in the contract.
5. checks if  $\mathcal{A}$  has deposited  $(\ddot{y} + \ddot{c}\ddot{h}) \cdot |P'|$  amount. If the check fails, it instructs the ledger to refund the coins that every party deposited and sends message  $\text{abort}_1$  to TTP (and accordingly to all parties); it outputs whatever  $\mathcal{A}$  outputs and then halts.

6. picks a random polynomial  $\zeta$  of degree 1, for each bin. Also,  $\text{Sim}_A^{\text{JUS}}$ , for each client  $C \in \{A_1, \dots, A_m\}$  allocates to each bin two random polynomials:  $(\omega^{(D,C)}, \rho^{(D,C)})$  of degree  $d$  and two random polynomials:  $(\gamma^{(D,C)}, \delta^{(D,C)})$  of degree  $3d+1$ . Moreover,  $\text{Sim}_A^{\text{JUS}}$  for every honest client  $C' \in \hat{P}$ , for each bin, picks two random polynomials:  $(\omega^{(C',D)}, \rho^{(C',D)})$  of degree  $d$ .
7. simulates  $\text{VOPR}$  using inputs  $\zeta \cdot \omega^{(D,C)}$  and  $\gamma^{(D,C)}$  for each bin. Accordingly, it receives the inputs of clients  $C'' \in P'$ , i.e.,  $\omega^{(C'',D)} \cdot \pi^{(C'')}$ , from its functionality  $f^{\text{VOPR}}$ , for each bin.
8. extracts the roots of polynomial  $\omega^{(C'',D)} \cdot \pi^{(C'')}$  for each bin and appends those roots that are in the sets universe to a new set  $S^{(C'')}$ .
9. simulates  $\text{VOPR}$  again using inputs  $\zeta \cdot \rho^{(D,C)} \cdot \pi^{(D)}$  and  $\delta^{(D,C)}$ , for each bin.
10. sends to TTP the input sets of all parties; namely, (i) client  $D$ 's input set:  $S^{(D)}$ , (ii) honest clients' input sets:  $S^{(C')}$  for all  $C'$  in  $\hat{P}$ , and (iii)  $\mathcal{A}$ 's input sets:  $S^{(C'')}$ , for all  $C''$  in  $P'$ . For each bin, it receives the intersection set,  $S_{\cap}$ , from TTP.
11. represents the intersection set for each bin as a polynomial,  $\pi$ , as follows:  

$$\pi = \prod_{i=1}^{|S_{\cap}|} (x - s_i), \text{ where } s_i \in S_{\cap}.$$
12. constructs polynomials  $\theta_1^{(C')} = \zeta \cdot \omega^{(D,C')} \cdot \omega^{(C',D)} \cdot \pi + \gamma^{(D,C')}$ ,  $\theta_2^{(C')} = \zeta \cdot \rho^{(D,C')} \cdot \rho^{(C',D)} \cdot \pi + \delta^{(D,C')}$ , and  $\nu^{(C')} = \theta_1^{(C')} + \theta_2^{(C')} + \tau^{(C')}$ , for each bin and each honest client  $C' \in \hat{P}$ , where  $\tau^{(C)} = \sum_{i=0}^{3d+2} z_{i,c} \cdot x^i$  and each  $z_{i,c}$  is derived from  $k$ .
13. sends to  $\mathcal{A}$  polynomial  $\nu^{(C')}$  for each bin and each client  $C'$ .
14. receives  $\nu^{(C'')}$  from  $\mathcal{A}$ , for each bin and each client  $C'' \in P'$ . It ensures that the output for every  $C''$  has been provided. Otherwise, it halts.
15. if there is any abort within steps 7–14, then it sends *abort*<sub>2</sub> to TTP and instructs the ledger to refund the coins that every party deposited. It outputs whatever  $\mathcal{A}$  outputs and then halts.
16. constructs polynomial  $\nu^{(D)} = \zeta \cdot \omega'^{(D)} \cdot \pi - \sum_{C=A_1}^{A_m} (\gamma^{(D,C)} + \delta^{(D,C)}) + \zeta \cdot \gamma'$  for each bin, where  $\omega'^{(D)}$  is a fresh random polynomial of degree  $d$  and  $\gamma'$  is a pseudorandom polynomial derived from  $mk$ .
17. sends to  $\mathcal{A}$  polynomials  $\nu^{(D)}$  and  $\zeta$  for each bin.
18. given each  $\nu^{(C'')}$ , computes polynomial  $\phi'$  as follows  $\phi' = \sum_{\forall C'' \in P'} \nu^{(C'')} - \sum_{\forall C'' \in P'} (\gamma^{(D,C'')} + \delta^{(D,C'')})$ , for every bin. Then,  $\text{Sim}_A^{\text{JUS}}$  checks whether  $\zeta$  divides  $\phi'$ , for every bin. If the check passes, it sets  $\text{Flag} = \text{True}$ . Otherwise, it sets  $\text{Flag} = \text{False}$ .
19. if  $\text{Flag} = \text{True}$ :
  - (a) instructs the ledger to send back each party's deposit, i.e.,  $\ddot{y} + \ddot{c}h$  amount. It sends a message *deliver* to TTP.
  - (b) outputs whatever  $\mathcal{A}$  outputs and then halts.
20. if  $\text{Flag} = \text{False}$ :
  - (a) receives  $|P'|$  keys of the PRF from  $\mathcal{A}$ , i.e.,  $\vec{k}' = [k'_1, \dots, k'_{|P'|}]$ , for every bin.

- (b) checks whether the following equation holds:  $k'_j = k$ , for every  $k'_j \in \vec{k}'$ . Note that  $k$  is the output of  $f^{\text{ZSPA-A}}$  generated in step 3. It constructs an empty list  $L'$  and appends to it the indices (e.g.,  $j$ ) of the keys that do not pass the above check.
- (c) simulates ZSPA-A and receives from  $f^{\text{ZSPA-A}}$  the output that contains a vector of random polynomials,  $\vec{\mu}'$ , for each valid key.
- (d) sends to  $\mathcal{A}$ , the list  $L'$  and vector  $\vec{\mu}'$ , for every bin.
- (e) for each bin of client  $C$  whose index (or ID) is not in  $L'$  computes polynomial  $\chi^{(D,C)}$  as follows:  $\chi^{(D,C)} = \zeta \cdot \eta^{(D,C)} - (\gamma^{(D,C)} + \delta^{(D,C)})$ , where  $\eta^{(D,C)}$  is a fresh random polynomial of degree  $3d + 1$ . Note,  $C$  includes both honest and corrupt clients, except those clients whose index is in  $L'$ .  $\text{Sim}_A^{\text{JUS}}$  sends every polynomial  $\chi^{(D,C)}$  to  $\mathcal{A}$ .
- (f) given each  $\nu^{(C'')}$  (by  $\mathcal{A}$  in step 14), computes polynomial  $\phi^{(C'')}$  as follows:  $\phi^{(C'')} = \nu^{(C'')} - \gamma^{(D,C'')} - \delta^{(D,C'')}$ , for every bin. Then,  $\text{Sim}_A^{\text{JUS}}$  checks whether  $\zeta$  divides  $\phi^{(C'')}$ , for every bin. It appends the index of those clients that did not pass the above check to a new list,  $L''$ . Note that  $L' \cap L'' = \perp$ .
- (g) if  $L'$  or  $L''$  is not empty, then it instructs the ledger: (i) to refund the coins of those parties whose index is not in  $L'$  and  $L''$ , (ii) to retrieve  $\check{c}h$  amount from the adversary (i.e., one of the parties whose index is in one of the lists) and send the  $\check{c}h$  amount to the auditor, and (iii) to compensate each honest party (whose index is not in the two lists)  $\frac{m' \cdot (\check{y} + \check{c}h) - \check{c}h}{m - m'}$  amount, where  $m' = |L'| + |L''|$ . Then, it sends message  $abort_3$  to TTP.
- (h) outputs whatever  $\mathcal{A}$  outputs and halts.

Next, we show that the real and ideal models are computationally indistinguishable. We first focus on the adversary's output. In the real and ideal models, the adversary sees the transcripts of ideal calls to  $f_{\text{ct}}$  as well as this functionality outputs, i.e.,  $mk$ . Due to the security of CT (as we are in the  $f_{\text{ct}}$ -hybrid world), the transcripts of  $f_{\text{ct}}$  in both models have identical distribution, so have the random output of  $f_{\text{ct}}$ , i.e.,  $mk$ . The same holds for (the transcripts and) outputs (i.e.,  $(k, g, q)$ ) of  $f^{\text{ZSPA-A}}$  that the adversary observes in the two models. Also, the deposit amount is identical in both models. Thus, in the case where  $abort_1$  is disseminated at this point; the adversary's output distribution in both models is identical.

The adversary also observes (the transcripts and) outputs of ideal calls to  $f^{\text{VOPR}}$  in both models, i.e., output  $(\theta_1^{(C'')} = \zeta \cdot \omega^{(D,C'')} \cdot \omega^{(C'',D)} \cdot \pi^{(C'')} + \gamma^{(D,C'')}, \theta_2^{(C'')} = \zeta \cdot \rho^{(D,C'')} \cdot \rho^{(C'',D)} \cdot \pi^{(D)} + \delta^{(D,C'')})$  for each corrupted client  $C''$ . However, due to the security of VOPR, the  $\mathcal{A}$ 's view, regarding VOPR, in both models have identical distribution. In the real model, the adversary observes the polynomial  $\nu^{(C)}$  that each honest client  $C$  stores in the smart contract. Nevertheless, this is a blinded polynomial comprising of two uniformly random blinding polynomials (i.e.,  $\gamma^{(D,C)}$  and  $\delta^{(D,C)}$ ) unknown to the adversary. In the ideal model,  $\mathcal{A}$  is given polynomial  $\nu^{(C')}$  for each honest client  $C'$ . This polynomial has also been blinded via two uniformly random polynomials (i.e.,  $\gamma^{(D,C')}$  and  $\delta^{(D,C')}$ ) unknown



to  $\mathcal{A}$ . Thus,  $\nu^{(C)}$  in the real model and  $\nu^{(C)}$  in the ideal model have identical distributions. As a result, in the case where  $abort_2$  is disseminated at this point; the adversary's output distribution in both models is identical.

Furthermore, in the real world, the adversary observes polynomials  $\zeta$  and  $\nu^{(D)}$  that  $D$  stores in the smart contract. Nevertheless,  $\zeta$  is a uniformly random polynomial, also polynomial  $\nu^{(D)}$  has been blinded; its blinding factors are the additive inverse of the sum of the random polynomials  $\gamma^{(D,C)}$  and  $\delta^{(D,C)}$  unknown to the adversary, for every client  $C \in \{A_1, \dots, A_m\}$  and  $D$ . In the ideal model,  $\mathcal{A}$  is given  $\zeta$  and  $\nu^{(D)}$ , where the former is a random polynomial and the latter is a blinded polynomial that has been blinded with the additive inverse of the sum of random polynomials  $\gamma^{(D,C)}$  and  $\delta^{(D,C)}$  unknown to it, for all client  $C$ . Therefore,  $(\zeta, \nu^{(D)})$  in the real model and  $(\zeta, \nu^{(D)})$  in the ideal model component-wise have identical distribution.

Also, the sum of less than  $m + 1$  blinded polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}$  in the real model has identical distribution to the sum of less than  $m + 1$  blinded polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}$  in the ideal model, as such a combination would still be blinded by a set of random blinding polynomials unknown to the adversary. Now we discuss why the two polynomials  $\frac{\phi}{\zeta} - \gamma'$  in the real model and  $\frac{\phi}{\zeta} - \gamma'$  in the ideal model are indistinguishable. Note that we divide and then subtract polynomials  $\phi$  because the adversary already knows (and must know) polynomials  $(\zeta, \gamma')$ . In the real model, polynomial  $\frac{\phi}{\zeta} - \gamma'$  has the following form:

$$\begin{aligned} \frac{\phi}{\zeta} - \gamma' &= \omega^{(D)} \cdot \pi^{(D)} + \sum_{C=A_1}^{A_m} (\omega^{(D,C)} \cdot \omega^{(C,D)} \cdot \pi^{(C)}) + \\ &\quad + \pi^{(D)} \cdot \sum_{C=A_1}^{A_m} (\rho^{(D,C)} \cdot \rho^{(C,D)}) \\ &= \mu \cdot gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)}) \end{aligned} \tag{3}$$

In Equation 3, every element of  $[\omega^{(D)}, \dots, \omega^{(D,C)}, \rho^{(D,C)}]$  is a uniformly random polynomial for every client  $C \in \{A_1, \dots, A_m\}$  (including corrupt ones) and client  $D$ ; because it has been picked by (in this case honest) client  $D$ . Thus, as shown in Section 3.9,  $\frac{\phi}{\zeta} - \gamma'$  has the form  $\mu \cdot gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)})$ , where  $\mu$  is a uniformly random polynomial and  $gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)})$  represents the intersection of the input sets.

In the ideal model,  $\mathcal{A}$  can construct polynomial  $\phi$  using its (well-formed) inputs  $\nu^{(C')}$  and polynomials  $\nu^{(C')}$  that the simulator has sent to it, for all  $C' \in \hat{P}$  and all  $C'' \in P'$ . Thus, in the ideal model, polynomial  $\frac{\phi}{\zeta} - \gamma'$  has the

following form:

$$\begin{aligned}
\frac{\phi}{\zeta} - \gamma' &= \pi \cdot \left( \sum_{\forall C' \in \bar{P}} (\omega^{(D, C')} \cdot \omega^{(C', D)}) + \sum_{\forall C' \in \bar{P}} (\rho^{(D, C')} \cdot \rho^{(C', D)}) \right) + \\
&+ \left( \sum_{\forall C'' \in P'} (\omega^{(D, C'')} \cdot \omega^{(C'', D)} \cdot \pi^{(C'')}) + \pi^{(D)} \cdot \sum_{\forall C'' \in P'} (\rho^{(D, C'')} \cdot \rho^{(C'', D)}) \right) \\
&= \mu \cdot \gcd(\pi, \pi^{(D)}, \pi^{(C'')})
\end{aligned} \tag{4}$$

In Equation 4, every element of the vector  $[\omega^{(D, C')}, \omega^{(D, C'')}, \rho^{(D, C')}, \rho^{(D, C'')}]$  is a uniformly random polynomial for all  $C' \in \bar{P}$  and all  $C'' \in P'$ , as they have been picked by  $\text{Sim}_A^{\text{JUS}}$ . Therefore,  $\frac{\phi}{\zeta} - \gamma'$  equals  $\mu \cdot \gcd(\pi, \pi^{(D)}, \pi^{(C'')})$ , such that  $\mu$  is a uniformly random polynomial and  $\gcd(\pi, \pi^{(D)}, \pi^{(C'')})$  represents the intersection of the input sets. We know that  $\gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)}) = \gcd(\pi, \pi^{(D)}, \pi^{(C'')})$ , as  $\pi$  includes the intersection of all clients' sets. Also,  $\mu$  has identical distribution in the two models, because they are uniformly random polynomials. Thus,  $\frac{\phi}{\zeta} - \gamma'$  in the real model and  $\frac{\phi}{\zeta} - \gamma'$  in the ideal model are indistinguishable.

Now we focus on the case where  $Flag = False$ . In the real model, the adversary observes the output of  $\text{Audit}(\cdot)$  which is a list of indices  $L$  and a vector of random polynomials  $\vec{\mu}$  picked by an honest auditor. In the ideal model,  $\mathcal{A}$  is given a list  $L'$  of indices and a vector of random polynomials  $\vec{\mu}'$  picked by the simulator. Thus, the pair  $(L, \vec{\mu})$  in the real model has identical distribution to the pair  $(L', \vec{\mu}')$  in the ideal model. Moreover, in the real model, the adversary observes each polynomial  $\chi^{(D, C)} = \zeta \cdot \eta^{(D, C)} - (\gamma^{(D, C)} + \delta^{(D, C)})$  that  $D$  stores in the contract, for each bin and each client  $C$  whose index is not in  $L$ . This is a blinded polynomial with blinding factor  $\eta^{(D, C)}$  which itself is a uniformly random polynomial picked by  $D$ . In the ideal model,  $\mathcal{A}$  is given a polynomial of the form  $\chi^{(D, C)} = \zeta \cdot \eta^{(D, C)} - (\gamma^{(D, C)} + \delta^{(D, C)})$ , for each bin and each client  $C$  whose index is not in  $L'$ . This is also a blinded polynomial whose blinding factor is  $\eta^{(D, C)}$  which itself is a random polynomial picked by the simulator. Therefore,  $\chi^{(D, C)}$  in the real model has identical distribution to  $\chi^{(D, C)}$  in the ideal model. In the real model, the adversary observes polynomial  $\iota^{(C)} = \zeta \cdot (\eta^{(D, C)} + \omega^{(D, C)} \cdot \omega^{(C, D)} \cdot \pi^{(C)} + \rho^{(D, C)} \cdot \rho^{(C, D)} \cdot \pi^{(D)} + \xi^{(C)})$  which is a blinded polynomial whose blinding factor is the sum of the above random polynomials, i.e.,  $\eta^{(D, C)} + \xi^{(C)}$ . In the ideal model,  $\mathcal{A}$  already has polynomials  $\chi^{(D, C)}, \nu^{(C)}$ , and  $\mu^{(C)}$ , where  $\mu^{(C)} \in \vec{\mu}'$ ; this lets  $\mathcal{A}$  compute  $\iota^{(C)} = \chi^{(D, C)} + \nu^{(C)} + \mu^{(C)} = \zeta \cdot (\eta^{(D, C)} + \omega^{(D, C')} \cdot \omega^{(C', D)} \cdot \pi + \rho^{(D, C')} \cdot \rho^{(C', D)} \cdot \pi + \xi^{(C)})$ , where  $\xi^{(C)}$  is a random blinding polynomial used in  $\mu^{(C)}$ . Nevertheless,  $\iota^{(C)}$  itself is a blinded polynomial whose blinding factor is the sum of random polynomials, i.e.,  $\eta^{(D, C)} + \xi^{(C)}$ . Hence, the distribution of polynomial  $\iota^{(C)}$  in the real model and  $\iota^{(C)}$  in the ideal model are identical. Moreover, the integer  $\ddot{y} + \ddot{c}h + \frac{m' \cdot (\ddot{y} + \ddot{c}h) - \ddot{c}h}{m - m'}$  has identical distribution in both models.

Next, we show that the honest party aborts with the same probability in the real and ideal models. Due to the security of CT, an honest party (during the execution of CT) aborts with the same probability in both models; in this case, the

adversary learns nothing about the parties' input set and the sets' intersection as the parties have not sent out any encoded input set yet. The same holds for the probability that honest parties abort during the execution of ZSPA-A. In this case, an aborting adversary also learns nothing about the parties' input set and the sets' intersection. Since all parties' deposit is public, an honest party can look up and detect if not all parties have deposited a sufficient amount with the same probability in both models. If parties halt because of insufficient deposit, no one learns about the parties' input set and the sets' intersection because the inputs (representation) have not been sent out at this point.

Due to the security of VOPR, honest parties abort with the same probability in both models. In the case of an abort during VOPR execution, the adversary would learn nothing (i) about its counter party' input set due to the security of VOPR, and (ii) about the rest of the honest parties' input sets and the intersection as the other parties' input sets are still blinded by random blinding factors known only to  $D$ . In the real model,  $D$  can check if all parties provided their encoded inputs, by reading from the smart contract. The simulator also can do the same check to ensure  $\mathcal{A}$  has provided the encoded inputs of all corrupt parties. Therefore, in both models, an honest party with the same probability would detect the violation of such a requirement, i.e., providing all encoded inputs. Even in this case, if an adversary aborts (by not proving its encoded inputs), then it learns nothing about the honest parties' input sets and the intersection for the reason explained above.

In the real model, the smart contract sums every client  $C$ 's polynomial  $\nu^{(C)}$  with each other and with  $D$ 's polynomial  $\nu^{(D)}$ , which removes the blinding factors that  $D$  initially inserted (during the execution of VOPR), and then checks whether the result is divisible by  $\zeta$ . Due to (a) Theorem 5 (i.e., unforgeable polynomials' linear combination), (b) the fact that the smart contract is given the random polynomial  $\zeta$  in plaintext, (c) no party (except honest client  $D$ ) knew anything about  $\zeta$  before they send their input to the contract, and (d) the security of the contract (i.e., the adversary cannot influence the correctness of the above verification performed by the contract), the contract can detect if a set of outputs of VOPR has been tampered with, with a probability at least  $1 - \epsilon(\lambda)$ . In the ideal model,  $\text{Sim}_A^{\text{JUS}}$  also can remove the blinding factors and it knows the random polynomial  $\zeta$ , unlike the adversary who does not know  $\zeta$  when it sends the outputs of VOPR to  $\text{Sim}_A^{\text{JUS}}$ . So,  $\text{Sim}_A^{\text{JUS}}$  can detect when  $\mathcal{A}$  modifies a set of the outputs of VOPR that were sent to  $\text{Sim}_A^{\text{JUS}}$  with a probability at least  $1 - \epsilon(\lambda)$ , due to Theorem 5. Hence, the smart contract in the real model and the simulator in the ideal model would abort with a similar probability.

Moreover, due to the security of ZSPA-A, the probability that an invalid key  $k_i \in \vec{k}$  is added to the list  $L$  in the real world is similar to the probability that  $\text{Sim}_A^{\text{JUS}}$  detects an invalid key  $k'_i \in \vec{k}'$  in the ideal world. In the real model, when  $\text{Flag} = \text{False}$ , the smart contract can identify each ill-structured output of VOPR (i.e.,  $\nu^{(C)}$ ) with a probability of at least  $1 - \epsilon(\lambda)$  by checking whether  $\zeta$  divides  $\nu^{(C)}$ , due to (a) Theorem 4 (i.e., unforgeable polynomial), (b) the fact that the smart contract is given  $\zeta$  in plaintext, (c) no party (except honest

client  $D$ ) knew anything about  $\zeta$  before they send their input to the contract, and (d) the security of the contract. In the ideal model, when  $Flag = False$ , given each  $\nu^{(c'')}$ ,  $\text{Sim}_A^{\text{JUS}}$  can remove its blinding factors from  $\nu^{(c'')}$  which results in  $\phi'^{(c'')}$  and then check if  $\zeta$  divides  $\phi'^{(c'')}$ . The simulator can also detect an ill-structured  $\nu^{(c'')}$  with a probability of at least  $1 - \epsilon(\lambda)$ , due to Theorem 4, the fact that the simulator is given  $\zeta$  in plaintext, and the adversary is not given any knowledge about  $\zeta$  before it sends to the simulator the outputs of VOPR. Hence, the smart contract in the real model and  $\text{Sim}_A^{\text{JUS}}$  in the ideal model would detect an ill-structured input of an adversary with the same probability.

Now, we analyse the output of the predicates ( $Q^{\text{Init}}, Q^{\text{Del}}, Q^{\text{UF-A}}, Q^{\text{F-A}}$ ) in the real and ideal models. In the real model, all clients proceed to prepare their input set only if the predefined amount of coins has been deposited by the parties; otherwise, they will be refunded and the protocol halts. In the ideal model also the simulator proceeds to prepare its inputs only if a sufficient amount of deposit has been put in the contract; otherwise, it would send message  $abort_1$  to TTP. Thus, in both models, the parties proceed to prepare their inputs only if  $Q^{\text{Init}}(.) \rightarrow 1$ . In the real model, if there is an abort after the parties ensure there is enough deposit and before client  $D$  provides its encoded input to the contract, then all parties would be able to retrieve their deposit in full; in this case, the aborting adversary would not be able to learn anything about honest parties input sets, because the parties' input sets are still blinded by random blinding polynomials known only to client  $D$ . In the ideal model, if there is any abort during steps 7–14, then the simulator sends  $abort_2$  to TTP and instructs the ledger to refund the coins that every party deposited. Also, in the case of an abort (within the above two points of time), the auditor is not involved. Thus, in both models, in the case of an abort within the above points of time, we would have  $Q^{\text{F-A}}(.) \rightarrow 1$ . In the real model, if  $Flag = True$ , then all parties would be able to learn the intersection and the smart contract refunds all parties, i.e., sends each party  $\ddot{y} + \ddot{c}h$  amount which is the amount each party initially deposited.

In the ideal model, when  $Flag = True$ , then  $\text{Sim}_A^{\text{JUS}}$  can extract the intersection (by summing the output of VOPR provided by all parties and removing the blinding polynomials) and sends back each party's deposit, i.e.,  $\ddot{y} + \ddot{c}h$  amount. Hence, in both models in the case of  $Flag = True$ , when all of the parties receive the result, we would have  $Q^{\text{Del}}(.) \rightarrow 1$ . In the real model, when  $Flag = False$ , only the adversary which might corrupt  $m'$  clients would be able to learn the result; in this case, the contract sends (i)  $\ddot{c}h$  amount to the auditor, and (ii)  $\frac{m' \cdot (\ddot{y} + \ddot{c}h) - \ddot{c}h}{m - m'}$  amount as a compensation, to each honest party, in addition to each party's deposit  $\ddot{y} + \ddot{c}h$ . In the ideal model, when  $Flag = False$ ,  $\text{Sim}_A^{\text{JUS}}$  sends  $abort_3$  to TTP and instructs the ledger to distribute the same amount among the auditor (e.g., with address  $adr_j$ ) and every honest party (e.g., with address  $adr_i$ ) as the contract does in the real model. Thus, in both models when  $Flag = False$ , we would have  $Q^{\text{UF-A}}(., ., ., ., adr_i) \rightarrow (a = 1, .)$  and  $Q^{\text{UF-A}}(., ., ., ., adr_j) \rightarrow (., b = 1)$ .

We conclude that the distributions of the joint outputs of the honest client  $C \in \hat{P}$ , client  $D$ ,  $Aud$ , and the adversary in the real and ideal models are computationally indistinguishable.

**Case 2: Corrupt dealer  $D$ .** In the real execution, the dealer's view is defined as follows:

$$\text{View}_D^{\text{JUS}}\left(S^{(D)}, (S^{(1)}, \dots, S^{(m)})\right) = \{S^{(D)}, \text{adr}_{sc}, m \cdot (\ddot{y} + \ddot{c}\ddot{h}), r_D, \text{View}_D^{\text{CT}}, k, g, q, \text{View}_D^{\text{VOPR}}, \nu^{(A_1)}, \dots, \nu^{(A_m)}, S_\cap\}$$

where  $\text{View}_D^{\text{CT}}$  and  $\text{View}_D^{\text{VOPR}}$  refer to the dealer's real-model view during the execution of CT and VOPR respectively. Also,  $r_D$  is the outcome of internal random coins of client  $D$  and  $\text{adr}_{sc}$  is the address of contract  $\mathcal{SC}_{\text{JUS}}$ . The simulator  $\text{Sim}_D^{\text{JUS}}$ , which receives all parties' input sets, works as follows.

1. receives from the subroutine adversary polynomials  $\zeta, (\gamma^{(A_1)}, \delta^{(A_1)}), \dots, (\gamma^{(A_m)}, \delta^{(A_m)}), (\omega^{(A_1)}, \rho^{(A_1)}), \dots, (\omega^{(A_m)}, \rho^{(A_m)})$ , where  $\deg(\gamma^{(C)}) = \deg(\delta^{(C)}) = 3d+1, \deg(\omega^{(C)}) = \deg(\rho^{(C)}) = d$ , and  $\deg(\zeta) = 1$ , and  $C \in \{A_1, \dots, A_m\}$ .
2. generates an empty view. It appends to the view, the input set  $S^{(D)}$ . It constructs and deploys a smart contract. Let  $\text{adr}_{sc}$  be the contract's address. It appends  $\text{adr}_{sc}$  to the view.
3. appends to the view integer  $m \cdot (\ddot{y} + \ddot{c}\ddot{h})$  and coins  $r'_D$  chosen uniformly at random.
4. extracts the simulation of CT from CT's simulator for client  $D$ . Let  $\text{Sim}_D^{\text{CT}}$  be the simulation. It appends  $\text{Sim}_D^{\text{CT}}$  to the view.
5. picks a random key,  $k'$ , and derives pseudorandom values  $z'_{i,j}$  from the key (the same way is done in Figure 2). It constructs a Merkle tree on top of all values  $z'_{i,j}$ . Let  $g'$  be the root of the resulting tree. It appends  $k', g'$ , and  $q' = H(k')$  to the view.
6. invokes VOPR's functionality twice and extracts the simulation of VOPR from VOPR's simulator for client  $D$ . Let  $\text{Sim}_D^{\text{VOPR}}$  be the simulation. It appends  $\text{Sim}_D^{\text{VOPR}}$  to the view.
7. given the parties' input sets, computes a polynomial  $\pi$  that represents the intersection of the sets.
8. picks  $m$  random polynomials  $\tau^{(A_1)}, \dots, \tau^{(A_m)}$  of degree  $3d+1$  such that their sum is 0.
9. picks  $m$  pairs of random polynomials  $(\omega^{(A_1)}, \rho^{(A_1)}), \dots, (\omega^{(A_m)}, \rho^{(A_m)})$ , where each polynomial is of degree  $d$ . Then,  $\text{Sim}_D^{\text{JUS}}$  for each client  $C \in \{A_1, \dots, A_m\}$  computes polynomial  $\nu^{(C)} = \zeta \cdot \pi \cdot (\omega^{(C)} \cdot \omega'^{(C)} + \rho^{(C)} \cdot \rho'^{(C)}) + \delta^{(C)} + \gamma^{(C)} + \tau^{(C)}$ .
10. appends  $\nu^{(A_1)}, \dots, \nu^{(A_m)}$  and the intersection of the sets  $S'_\cap$  to the view.

Next, we will show that the two views are computationally indistinguishable.  $D$ 's input  $S^{(D)}$  is identical in both models; therefore, they have identical distributions. Also, the contract's address has the same distribution in both views, and so has the integer  $m \cdot (\ddot{y} + \ddot{c}\ddot{h})$ . Since the real-model semi-honest adversary

samples its randomness according to the protocol's description, the random coins in both models (i.e.,  $r_D$  and  $r'_D$ ) have identical distributions. Moreover, due to the security of CT,  $\text{View}_D^{\text{CT}}$  and  $\text{Sim}_D^{\text{CT}}$  have identical distributions. Keys  $k$  and  $k'$  have identical distributions, as both have been picked uniformly at random from the same domain. In the real model, each element of the pair  $(g, p)$  is the output of a deterministic function on a random value  $k$ . We know that  $k$  in the real model has identical distribution to  $k'$  in the ideal model, so do the evaluations of deterministic functions (i.e., Merkle tree, H, and PRF) on them. Therefore, each pair  $(g, q)$  in the real model component-wise has an identical distribution to each pair  $(g', q')$  in the ideal model. Furthermore, due to the security of the VOPR,  $\text{View}_D^{\text{VOPR}}$  and  $\text{Sim}_D^{\text{VOPR}}$  have identical distributions.

In the real model, each  $\nu^{(C)}$  has been blinded by a pseudorandom polynomial (i.e., derived from PRF's output) unknown to client  $D$ . In the ideal model, however, each  $\nu^{(C)}$  has been blinded by a random polynomial unknown to client  $D$ . Due to the security of PRF, its outputs are computationally indistinguishable from truly random values. Therefore,  $\nu^{(C)}$  in the real model and  $\nu^{(C)}$  in the ideal model are computationally indistinguishable. Now we focus on the sum of all  $\nu^{(C)}$  in the real model and the sum of all  $\nu^{(C)}$  in the ideal model, as adding them together would remove the above blinding polynomials that are unknown to client  $D$ . Specifically, in the real model, after client  $D$  sums all  $\nu^{(C)}$  and removes the blinding factors and  $\zeta$  that it initially imposed, it would get a polynomial of

$$\text{the form } \hat{\phi} = \frac{\sum_{C=A_1}^{A_m} \nu^{(C)} - \sum_{C=A_1}^{A_m} (\gamma^{(D,C)} + \delta^{(D,C)})}{\zeta} = \sum_{C=A_1}^{A_m} (\omega^{(D,C)} \cdot \omega^{(C,D)} \cdot \pi^{(C)}) + \pi^{(D)} \cdot$$

$\sum_{C=A_1}^{A_m} (\rho^{(D,C)} \cdot \rho^{(C,D)})$ , where  $\omega^{(C,D)}$  and  $\rho^{(C,D)}$  are random polynomials unknown to client  $D$ . In the ideal model, after summing all  $\nu^{(C)}$  and removing the random polynomials that it already knows, it would get a polynomial of the following

$$\text{form: } \hat{\phi}' = \frac{\sum_{C=A_1}^{A_m} \nu^{(C)} - \sum_{C=A_1}^{A_m} (\gamma^{(C)} + \delta^{(C)})}{\zeta} = \pi \cdot \sum_{C=A_1}^{A_m} (\omega'^{(C)} \cdot \omega^{(C)}) + (\rho'^{(C)} \cdot \rho^{(C)}).$$

As shown in Section 3.9, polynomial  $\hat{\phi}$  has the form  $\mu \cdot \gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)})$ , where  $\mu$  is a uniformly random polynomial and  $\gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)})$  represents the intersection of the input sets. Moreover, it is evident that  $\hat{\phi}'$  has the form  $\mu \cdot \pi$ , where  $\mu$  is a random polynomial and  $\pi$  represents the intersection. We know that both  $\gcd(\pi^{(D)}, \pi^{(A_1)}, \dots, \pi^{(A_m)})$  and  $\pi$  represent the same intersection, also  $\mu$  in the real model and  $\mu$  in the ideal model have identical distribution as they are uniformly random polynomials. Thus, two polynomials  $\hat{\phi}$  and  $\hat{\phi}'$  are indistinguishable. Also, the output  $S_{\cap}$  is identical in both views. We conclude that the two views are computationally indistinguishable.

**Case 3: Corrupt auditor.** In this case, by using the proof that we have already provided for Case 1 (i.e.,  $m - 1$  client  $A_j$ s are corrupt), we can easily construct a simulator that generates a view computationally distinguishable from the real-model semi-honest auditor. The reason is that, in the worst-case sce-

nario where  $m - 1$  malicious client  $A_j$ s reveal their input sets and randomness to the auditor, the auditor's view would be similar to the view of these corrupt clients, which we have shown to be indistinguishable. The only extra messages the auditor generates, that a corrupt client  $A_j$  would not see in plaintext, are random blinding polynomials  $(\xi^{(A_1)}, \dots, \xi^{(A_m)})$  generated during the execution of  $\text{Audit}(\cdot)$  of ZSPA-A; however, these polynomials are picked uniformly at random and independent of the parties' input sets. Thus, if the smart contract detects misbehaviour and invokes the auditor, even if  $m - 1$  corrupt client  $A_j$  reveals their input sets, then the auditor cannot learn anything about honest parties' input sets.

**Case 4: Corrupt public.** In the real model, the view of the public (i.e., non-participants of the protocol) is defined as below:

$$\text{View}_{Pub}^{\text{JUS}}(\perp, S^{(D)}, (S^{(A_1)}, \dots, S^{(A_m)})) = \{\perp, \text{adr}_{sc}, (m+1) \cdot (\ddot{y} + \ddot{c}h), k, g, q, \nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}\}$$

Now, we describe how the simulator  $\text{Sim}_{Pub}^{\text{JUS}}$  works.

1. generates an empty view and appends to it an empty symbol,  $\perp$ . It constructs and deploys a smart contract. It appends the contract's address,  $\text{adr}_{sc}$  and integer  $(m+1) \cdot (\ddot{y} + \ddot{c}h)$  to the view.
2. picks a random key,  $k'$ , and derives pseudorandom values  $z'_{i,j}$  from the key, in the same way, done in Figure 2. It constructs a Merkle tree on top of the  $z'_{i,j}$  values. Let  $g'$  be the root of the resulting tree. It appends  $k', g'$ , and  $q' = H(k')$  to the view.
3. for each client  $C \in \{A_1, \dots, A_m\}$  and client  $D$  generates a random polynomial of degree  $3d+1$  (for each bin), i.e.,  $\nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}$ .

Next, we will show that the two views are computationally indistinguishable. In both views,  $\perp$  is identical. Also, the contract's addresses (i.e.,  $\text{adr}_{sc}$ ) has the same distribution in both views, and so has the integer  $(m+1) \cdot (\ddot{y} + \ddot{c}h)$ . Keys  $k$  and  $k'$  have identical distributions as well, because both of them have been picked uniformly at random from the same domain. In the real model, each element of pair  $(g, p)$  is the output of a deterministic function on the random key  $k$ . We know that  $k$  in the real model has identical distribution to  $k'$  in the ideal model, and so do the evaluations of deterministic functions on them. Hence, each pair  $(g, q)$  in the real model component-wise has an identical distribution to each pair  $(g', q')$  in the ideal model. In the real model, each polynomial  $\nu^{(C)}$  is a blinded polynomial comprising of two uniformly random blinding polynomials (i.e.,  $\gamma^{(D,C)}$  and  $\delta^{(D,C)}$ ) unknown to the adversary. In the ideal model, each polynomial  $\nu^{(C)}$  is a random polynomial; thus, polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}$  in the real model have identical distribution to polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}$  in the ideal model. Similarly, polynomial  $\nu^{(D)}$  has been blinded in the real model; its blinding factors are the additive inverse of the sum of the random polynomials  $\gamma^{(D,C)}$  and  $\delta^{(D,C)}$  unknown to the adversary. In the ideal model, polynomial  $\nu^{(D)}$

is a uniformly random polynomial; thus,  $\nu^{(D)}$  in the real model and  $\nu^{(D)}$  in the ideal model have identical distributions. Moreover, in the real model even though the sum  $\phi$  of polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}$  would remove some of the blinding random polynomials, it is still a blinded polynomial with a pseudorandom blinding factor  $\gamma'$  (derived from the output of PRF), unknown to the adversary. In the ideal model, the sum of polynomials  $\nu^{(A_1)}, \dots, \nu^{(A_m)}, \nu^{(D)}$  is also a random polynomial. Thus, the sum of the above polynomials in the real model is computationally indistinguishable from the sum of those polynomials in the ideal model. We conclude that the two views are computationally indistinguishable.

## P Formal Definitions of Predicates of $\mathcal{PSI}^{\mathcal{FCR}}$

Below, we present the formal definition of predicates  $Q_R^{\text{Del}}$  and  $Q_R^{\text{UF-A}}$ .

**Definition 11 ( $Q_R^{\text{Del}}$ : Delivery-with-Reward predicate).** *Let  $\mathcal{G}$  be a stable ledger,  $\text{adr}_{sc}$  be smart contract sc's address,  $\text{adr}_i \in \text{Adr}$  be the address of an honest party,  $\ddot{x}$  be a fixed amount of coins, and  $\text{pram} := (\mathcal{G}, \text{adr}_{sc}, \ddot{x})$ . Let  $R$  be a reward function that takes as input the computation result:  $\text{res}$ , a party's address:  $\text{adr}_i$ , a reward a party should receive for each unit of revealed information:  $\ddot{l}$ , and input size:  $\text{inSize}$ . Then  $R$  is defined as follows, if  $\text{adr}_i$  belongs to a non-buyer, then it returns the total amount that  $\text{adr}_i$  should be rewarded and if  $\text{adr}_i$  belongs to a buyer client, then it returns the reward's leftover that the buyer can collect, i.e.,  $R(\text{res}, \text{adr}_i, \ddot{l}, \text{inSize}) \rightarrow \text{r}\ddot{w}_i$ . Then, the delivery with reward predicate  $Q_R^{\text{Del}}(\text{pram}, \text{adr}_i, \text{res}, \ddot{l}, \text{inSize})$  returns 1 if  $\text{adr}_i$  has sent  $\ddot{x}$  amount to  $sc$  and received at least  $\ddot{x} + \text{r}\ddot{w}_i$  amount from it. Else, it returns 0.*

**Definition 12 ( $Q_R^{\text{UF-A}}$ : UnFair-Abort-with-Reward predicate).** *Let  $\text{pram} := (\mathcal{G}, \text{adr}_{sc}, \ddot{x})$  be the parameters defined above, and  $\text{Adr}' \subset \text{Adr}$  be a set containing honest parties' addresses,  $m' = |\text{Adr}'|$ , and  $\text{adr}_i \in \text{Adr}'$ . Let also  $G$  be a compensation function that takes as input three parameters  $(\text{d}\ddot{e}\text{p}\text{s}, \text{adr}_i, m')$ , where  $\text{d}\ddot{e}\text{p}\text{s}$  is the amount of coins that all  $m + 1$  parties deposit,  $\text{adr}_i$  is an honest party's address, and  $m' = |\text{Adr}'|$ ; it returns the amount of compensation each honest party must receive, i.e.,  $G(\text{d}\ddot{e}\text{p}\text{s}, \text{adr}_i, m') \rightarrow \ddot{x}_i$ . Let  $R$  be the reward function defined above, i.e.,  $R(\text{res}, \text{adr}_i, \ddot{l}, \text{inSize}) \rightarrow \text{r}\ddot{w}_i$ , and let  $\text{pr}\ddot{a}\text{m} := (\text{res}, \ddot{l}, \text{inSize})$ . Then, predicate  $Q_R^{\text{UF-A}}$  is defined as  $Q_R^{\text{UF-A}}(\text{pram}, \text{pr}\ddot{a}\text{m}, G, R, \text{d}\ddot{e}\text{p}\text{s}, m', \text{adr}_i) \rightarrow (a, b)$ , where  $a = 1$  if  $\text{adr}_i$  is an honest party's address which has sent  $\ddot{x}$  amount to  $sc$  and received  $\ddot{x} + \ddot{x}_i + \text{r}\ddot{w}_i$  from it, and  $b = 1$  if  $\text{adr}_i$  is an auditor's address which received  $\ddot{x}_i$  from  $sc$ . Otherwise,  $a = b = 0$ .*

## Q Workflow of Anesidora

Figure 7 outlines the interaction between parties in Anesidora.



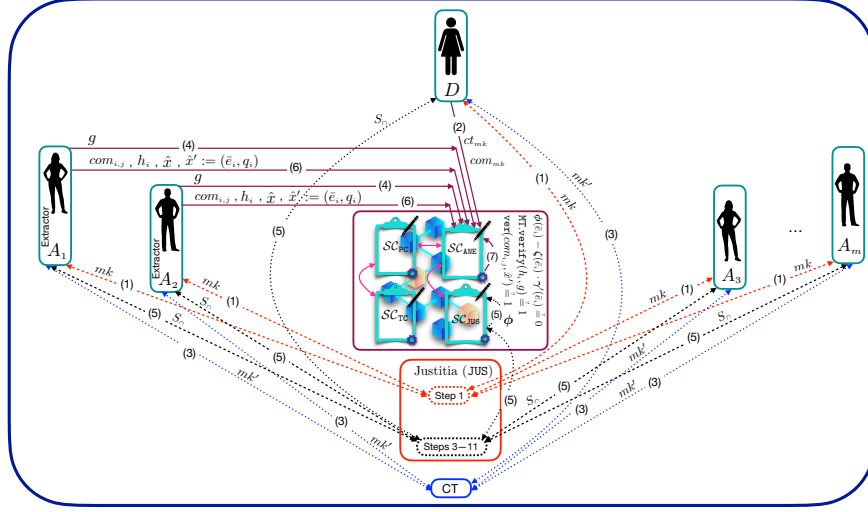


Fig. 7: Outline of the interactions between parties in Anesidora

## R Main Challenges that Anesidora Overcomes

### R.1 Rewarding Clients Proportionate to the Intersection Size.

In PSIs, the main private information about the clients which is revealed to a result recipient is the private set elements that the clients have in common. Thus, honest clients must receive a reward proportionate to the intersection cardinality, from a buyer. To receive the reward, the clients need to reach a consensus on the intersection cardinality. The naive way to do that is to let every client find the intersection and declare it to the smart contract. Under the assumption that the majority of clients are honest, then the smart contract can reward the honest result recipient (from the buyer's deposit). But, the honest majority assumption is strong in the context of multi-party PSI. Moreover, this approach requires all clients to extract the intersection, which would increase the overall costs. Some clients may not even be interested in or available to do so. This task could also be conducted by a single entity, such as the dealer; but this approach would introduce a single point of failure and all clients have to depend on this entity. To address these challenges, we allow any two clients to become extractors. Each of them finds and sends to the contract the (encrypted) elements in the intersection. It is paid by the contract if the contract concludes that it is honest. This lets us avoid (i) the honest majority assumption, (ii) requiring all clients to find the intersection, and (iii) relying on a single trusted/semi-honest party to complete the task.

## R.2 Dealing with Extractors' Collusion.

Using two extractors itself introduces another challenge; namely, they may collude with each other (and the buyer) to provide a consistent but incorrect result. This behaviour can not be detected by a verifier unless it always conducts the delegated task itself, which would defeat the purpose of delegation. To efficiently address this challenge, we use the counter-collusion smart contracts (from Section 3.4) that creates distrust among the extractors and incentivises them to act honestly.

## S Proof of ANE

In this section, we prove Theorem 7, i.e., the security of ANE.

*Proof.* We prove the theorem by considering the case where each party is corrupt, at a time.

**Case 1: Corrupt extractors  $\{A_1, A_2\}$  and  $m - 3$  clients in  $\{A_3, \dots, A_m\}$ .** Let set  $G$  include extractors  $\{A_1, A_2\}$  and a set of at most  $m - 3$  corrupt clients in  $\{A_3, \dots, A_m\}$ . Let set  $\hat{G}$  be a set of honest clients in  $\{A_3, \dots, A_m\}$ . Also, let  $\text{Sim}_A^{\text{ANE}}$  be the simulator. We let the simulator interact with (i) active adversary  $\mathcal{A}'$  that may corrupt  $m - 3$  clients in  $\{A_3, \dots, A_m\}$ , and (ii) two rational adversaries  $\mathcal{A}'' := (A_1, A_2)$  that corrupt extractors  $(A_1, A_2)$  component-wise. In the simulation, before the point where the extractors are invoked to provide proofs and results, the simulator directly deals with active adversary  $\mathcal{A}'$ . However, when the extractors are involved (to generate proofs and extract the result) we require the simulator to interact with each rational adversary  $A_1$  and  $A_2$ . We allow these two subroutine adversaries  $(\mathcal{A}', \mathcal{A}'')$  to internally interact with each other. Now, we explain how  $\text{Sim}_A^{\text{ANE}}$ , which receives the input sets of honest dealer  $D$  and honest client(s) in  $\hat{G}$ , works.

1. constructs and deploys two smart contracts (for JUS and ANE). It sends the contracts' addresses to  $\mathcal{A}'$ . It also simulates CT and receives the output value,  $mk$ , from its functionality,  $f_{\text{CT}}$ .
2. deposits  $S_{\min} \cdot \ddot{v}$  amount to  $\mathcal{SC}_{\text{ANE}}$  if buyer  $A_m$  is honest, i.e.,  $A_m \in \hat{G}$ . Otherwise,  $\text{Sim}_A^{\text{ANE}}$  checks if  $\mathcal{A}'$  has deposited  $S_{\min} \cdot \ddot{v}$  amount in  $\mathcal{SC}_{\text{ANE}}$ . If the check fails, it instructs the ledger to refund the coins that every party deposited and sends message  $\text{abort}_1$  to TTP (and accordingly to all parties); it outputs whatever  $\mathcal{A}'$  outputs and then halts.
3. constructs and deploys a (Prisoner's) contract and transfers  $\ddot{w} = S_{\min} \cdot \ddot{r}$  amount for each extractor.  $\text{Sim}_A^{\text{ANE}}$  ensures that each extractor deposited  $\ddot{d}' = \ddot{d} + S_{\min} \cdot \ddot{f}$  coins in this contract; otherwise, it instructs the ledger to refund the coins that every party deposited and sends message  $\text{abort}_1$  to TTP; it outputs whatever  $\mathcal{A}'$  outputs and then halts.
4. encrypts  $mk$  under the public key of the dispute resolver; let  $ct_{mk}$  be the resulting ciphertext. It also generates a commitment of  $mk$  as follows:  $com_{mk} = \text{Com}(mk, \text{PRF}(mk, 0))$ . It stores  $ct_{mk}$  and  $com_{mk}$  in  $\mathcal{SC}_{\text{ANE}}$ .

5. simulates CT again and receives the output value  $mk'$  from  $f_{\text{CT}}$ .
6. receives from  $\mathcal{A}'$  a Merkle tree's root  $g'$  for each extractor.
7. simulates the steps of 3–11 in JUS. For completeness, we include the steps that the simulator takes in this proof. Specifically,  $\text{Sim}_A^{\text{ANE}}$ :
  - (a) simulates ZSPA-A for each bin and receives the output value  $(k, g, q)$  from  $f_{\text{ZSPA-A}}$ .
  - (b) deposits in the contract the amount of  $\ddot{y}' = S_{\text{min}} \cdot \ddot{v} + \ddot{c}h$  for client  $D$  and each honest client in  $\hat{G}$ . It sends to  $\mathcal{A}'$  the amount deposited in the contract.
  - (c) checks if  $\mathcal{A}'$  has deposited  $\ddot{y}' \cdot |G|$  amount (in addition to  $\ddot{d}'$  amount deposited in step 3 above). If the check fails, it instructs the ledger to refund the coins that every party deposited and sends message  $\text{abort}_1$  to TTP (and accordingly to all parties); it outputs whatever  $\mathcal{A}'$  outputs and then halts.
  - (d) picks a random polynomial  $\zeta$  of degree 1, for each bin.  $\text{Sim}_A^{\text{ANE}}$ , for each client  $C \in \{A_1, \dots, A_m\}$  allocates to each bin two degree  $d$  random polynomials:  $(\omega^{(D,C)}, \rho^{(D,C)})$ , and two degree  $3d + 1$  random polynomials:  $(\gamma^{(D,C)}, \delta^{(D,C)})$ . Also,  $\text{Sim}_A^{\text{ANE}}$  for each honest client  $C' \in \hat{G}$ , for each bin, picks two degree  $d$  random polynomials:  $(\omega^{(C',D)}, \rho^{(C',D)})$ .
  - (e) simulates VOPR using inputs  $\zeta \cdot \omega^{(D,C)}$  and  $\gamma^{(D,C)}$  for each bin. It receives the inputs of clients  $C'' \in G$ , i.e.,  $\omega^{(C'',D)} \cdot \pi^{(C'')}$ , from its functionality  $f^{\text{VOPR}}$ , for each bin.
  - (f) extracts the roots of polynomial  $\omega^{(C'',D)} \cdot \pi^{(C'')}$  for each bin and appends those roots that are in the sets universe to a new set  $S^{(C'')}$ .
  - (g) simulates again VOPR using inputs  $\zeta \cdot \rho^{(D,C)} \cdot \pi^{(D)}$  and  $\delta^{(D,C)}$ , for each bin.
  - (h) sends to TTP the input sets of all parties; specifically, (i) client  $D$ 's input set:  $S^{(D)}$ , (ii) honest clients' input sets:  $S^{(C')}$  for all  $C'$  in  $\hat{G}$ , and (iii)  $\mathcal{A}'$ 's input sets:  $S^{(C'')}$ , for all  $C''$  in  $G$ . For each bin, it receives the intersection set,  $S_{\cap}$ , from TTP.
  - (i) represents the intersection set for each bin as a polynomial,  $\pi$ , as follows. First, it encrypts each element  $s_i$  of  $S_{\cap}$  as  $e_i = \text{PRP}(mk', s_i)$ . Second, it encodes each encrypted element as  $\bar{e}_i = e_i || H(e_i)$ . Third, it constructs  $\pi$  as  $\pi = \prod_{i=1}^{|S_{\cap}|} (x - s_i) \cdot \prod_{j=1}^{d-|S_{\cap}|} (x - u_j)$ , where  $u_j$  is a dummy value.
  - (j) constructs polynomials  $\theta_1^{(C')} = \zeta \cdot \omega^{(D,C')} \cdot \omega^{(C',D)} \cdot \pi + \gamma^{(D,C')}$ ,  $\theta_2^{(C')} = \zeta \cdot \rho^{(D,C')} \cdot \rho^{(C',D)} \cdot \pi + \delta^{(D,C')}$ , and  $\nu^{(C')} = \theta_1^{(C')} + \theta_2^{(C')} + \tau^{(C')}$ , for each bin and each honest client  $C' \in \hat{G}$ , such that  $\tau^{(C)} = \sum_{i=0}^{3d+2} z_{i,c} \cdot x^i$  and each value  $z_{i,c}$  is derived from key  $k$  generated in step 7a. It sends to  $\mathcal{A}'$  polynomial  $\nu^{(C')}$  for each bin and each honest client  $C' \in \hat{G}$ .
  - (k) receives  $\nu^{(C'')}$  from  $\mathcal{A}'$ , for each bin and each corrupt client  $C'' \in G$ . It checks whether the output for every  $C''$  has been provided. Otherwise, it halts.
  - (l) if there is any abort within steps 7e–7k, then it sends  $\text{abort}_2$  to TTP and instructs the ledger to refund the coins that every party deposited. It outputs whatever  $\mathcal{A}'$  outputs and then halts.

- (m) constructs polynomial  $\nu^{(D)} = \zeta \cdot \omega'^{(D)} \cdot \pi - \sum_{C=A_1}^{A_m} (\gamma^{(D,C)} + \delta^{(D,C)}) + \zeta \cdot \gamma'$  for each bin on behalf of client  $D$ , where  $\omega'^{(D)}$  is a fresh random polynomial of degree  $d$  and  $\gamma'$  is a pseudorandom polynomial derived from  $mk$ .
- (n) sends to  $\mathcal{A}'$  polynomials  $\nu^{(D)}$  and  $\zeta$  for each bin.
- (o) computes polynomial  $\phi'$  as  $\phi' = \sum_{\forall C'' \in G} \nu^{(C'')} - \sum_{\forall C'' \in G} (\gamma^{(D,C'')} + \delta^{(D,C'')})$ , for every bin. Next, it checks if  $\zeta$  divides  $\phi'$ , for every bin. If the check passes, it sets  $Flag = True$ . Otherwise, it sets  $Flag = False$ .
- (p) if  $Flag = True$ , then instructs the ledger to send back each party's deposit, i.e.,  $\ddot{y}'$  amount. It sends a message *deliver* to TTP. It proceeds to step 8 below.
- (q) if  $Flag = False$ :
- i. receives  $|G|$  keys of the PRF from  $\mathcal{A}'$ , i.e.,  $\vec{k}' = [k'_1, \dots, k'_{|G|}]$ , for every bin.
  - ii. checks if  $k'_j = k$ , for every  $k'_j \in \vec{k}'$ . Recall,  $k$  was generated in step 3. It constructs an empty list  $L'$  and appends to it the indices (e.g.,  $j$ ) of the keys that do not pass the above check.
  - iii. receives from  $f^{\text{ZSPA-A}}$  the output containing a vector of random polynomials,  $\vec{\mu}'$ , for each valid key.
  - iv. sends to  $\mathcal{A}'$ ,  $L'$  and  $\vec{\mu}'$ , for every bin.
  - v. for each bin of client  $C$  whose index is not in  $L'$  computes polynomial  $\chi^{(D,C)}$  as  $\chi^{(D,C)} = \zeta \cdot \eta^{(D,C)} - (\gamma^{(D,C)} + \delta^{(D,C)})$ , where  $\eta^{(D,C)}$  is a fresh random polynomial of degree  $3d + 1$ . Note that  $C$  includes both honest and corrupt clients, except those clients whose index is in  $L'$ .  $\text{Sim}_A^{\text{ANE}}$  sends every polynomial  $\chi^{(D,C)}$  to  $\mathcal{A}'$ .
  - vi. given each  $\nu^{(C'')}$  (by  $\mathcal{A}'$  in step 7k), computes polynomial  $\phi'^{(C'')}$  as follows:  $\phi'^{(C'')} = \nu^{(C'')} - \gamma^{(D,C'')} - \delta^{(D,C'')}$ , for every bin.  $\text{Sim}_A^{\text{ANE}}$  checks if  $\zeta$  divides  $\phi'^{(C'')}$ , for every bin. It appends the index of those clients that did not pass the above check to a new list,  $L''$ .
  - vii. if  $L'$  or  $L''$  is not empty, then instructs the ledger: (a) to refund  $\ddot{y}'$  amount to each client whose index is not in  $L'$  and  $L''$ , (b) to retrieve  $\ddot{c}h$  amount from the adversary (i.e., one of the parties whose index is in one of the lists) and send the  $\ddot{c}h$  amount to *Aud*, and (c) to reward and compensate each honest party (whose index is not in the two lists)  $\frac{m' \cdot \ddot{y}' - \ddot{c}h}{m - m'}$  amount, where  $m' = |L'| + |L''|$ . Then, it sends message *abort*<sub>3</sub> to TTP.
  - viii. outputs whatever  $\mathcal{A}'$  outputs and halts.
8. for each  $I \in \{1, 2\}$ , receives from  $\mathcal{A}_I$  (1) a set  $E^{(I)}$  of encoded encrypted elements, e.g.,  $\bar{e}_i$ , in the intersection, (2) each  $\bar{e}_i$ 's commitment  $com_{i,j}$ , (3) each  $com_{i,j}$ 's opening  $\hat{x}'$ , (4) a proof  $h_i$  that each  $com_{i,j}$  is a leaf node of a Merkle tree with root  $g'$  (given to simulator in step 6 above), and (5) the opening  $\hat{x}$  of commitment  $com_{mk}$ .
9. encrypts each element  $s_i$  of  $S_\cap$  as  $e_i = \text{PRP}(mk', s_i)$ . Then, it encodes each encrypted element as  $\bar{e}_i = e_i \parallel \mathbb{H}(e_i)$ . Let set  $S'$  include all encoded encrypted elements in the intersection.

10. sings a  $\mathcal{SC}_{\text{TC}}$  with  $\mathcal{A}_I$ , if  $\mathcal{A}_I$  decides to be a traitor extractor. In this case,  $\mathcal{A}_I$ , provides the intersection to  $\mathcal{SC}_{\text{TC}}$ .  $\text{Sim}_A^{\text{ANE}}$  checks this intersection's validity. Shortly (in step 16b), we will explain how  $\text{Sim}_A^{\text{ANE}}$  acts based on the outcome of this check.
11. checks if each set  $E^{(I)}$  equals set  $S'$ .
12. checks if  $\text{com}_{i,j}$  matches the opening  $\hat{x}'$  and the opening corresponds to a unique element in  $S'$ .
13. verifies each commitment's proof,  $h_i$ . Specifically, given the proof and root  $g$ , it ensures the commitment  $\text{com}_{i,j}$  is a leaf node of a Merkle tree with a root node  $g'$ . It also checks whether the opening  $\hat{x}$  matches  $\text{com}_{mk}$ .
14. if all the checks in steps 11–13 pass, then instructs the ledger (i) to take  $|S_{\cap}| \cdot m \cdot \check{l}$  amount from the buyer's deposit and distributes it among all clients, except the buyer, (ii) to return the extractors deposit (i.e.,  $\check{d}'$  amount each) and pay each extractor  $|S_{\cap}| \cdot \check{r}$  amount, and (iii) to return  $(S_{\min} - |S_{\cap}|) \cdot \check{v}$  amount to the buyer.
15. if neither extractor sends the extractor's set intersection  $(E^{(A_1)}, E^{(A_2)})$  in step 8, then instructs the ledger (i) to refund the buyer, by sending  $S_{\min} \cdot \check{v}$  amount back to the buyer and (ii) to retrieve each extractor's deposit (i.e.,  $\check{d}'$  amount) from  $\mathcal{SC}_{\text{PC}}$  and distribute it among the rest of the clients (except the buyer and extractors).
16. if the checks in step 14 fail or in step 15 both  $\mathcal{A}_1$  and  $\mathcal{A}_2$  send the extractors' set intersection but they are inconsistent with each other, then it tags the extractor whose proof or set intersection was invalid as a misbehaving extractor.  $\text{Sim}_A^{\text{ANE}}$  instructs the ledger to pay the auditor (of  $\mathcal{SC}_{\text{PC}}$ ) the total amount of  $\check{c}h$  coins taken from the misbehaving extractor(s) deposit. Furthermore,  $\text{Sim}_A^{\text{ANE}}$  takes the following steps.
  - (a) if both extractors cheated and there is no traitor, then instructs the ledger (i) to refund the buyer  $S_{\min} \cdot \check{v}$  amount, and (ii) to take  $2\check{d}' - \check{c}h$  amount from the misbehaving extractors' deposit and distribute it to the rest of the clients except the buyer and extractors.
  - (b) if both extractors cheated, there is a traitor, and the traitor delivered a correct result (in step 10), then instructs the ledger (i) to take  $\check{d}' - \check{d}$  amount from the other misbehaving extractor's deposit and distribute it among the rest of the clients (except the buyer and dishonest extractor), (ii) to distribute  $|S_{\cap}| \cdot \check{r} + \check{d}' + \check{d} - \check{c}h$  amount to the traitor, (iii) to refund the traitor  $\check{c}h$  amount, and (iv) to refund the buyer  $S_{\min} \cdot \check{v} - |S_{\cap}| \cdot \check{r}$  amount.
  - (c) if both extractors cheated, there is a traitor, and the traitor delivered an incorrect result (in step 10), then instructs the ledger to distribute coins the same way it does in step 16a.
  - (d) if one of the extractors cheated and there is no traitor, then instructs the ledger (i) to return the honest extractor's deposit (i.e.,  $\check{d}'$  amount), (ii) to pay the honest extractor  $|S_{\cap}| \cdot \check{r}$  amount, (iii) to pay this extractor  $\check{d} - \check{c}h$  amount taken from the dishonest extractor's deposit, and (iv) to pay the buyer and the rest of the clients the same way it does in step 16b.

- (e) if one of the extractors cheated, there is a traitor, and the traitor delivered a correct result (in step 10), then instructs the ledger (i) to return the other honest extractor's deposit (i.e.,  $\ddot{d}'$  amount), (ii) to pay the honest extractor  $|S_\cap| \cdot \ddot{r}$  amount, taken from the buyer's deposit, (iii) to pay the honest extractor  $\ddot{d} - \ddot{c}h$  amount, taken from the traitor's deposit, (iv) to pay to the traitor  $|S_\cap| \cdot \ddot{r}$  amount, taken from the buyer's deposit, (v) to refund the traitor  $\ddot{d}' - \ddot{d}$  amount, (vi) to refund the traitor  $\ddot{c}h$  amount, (vii) to take  $|S_\cap| \cdot m \cdot \ddot{l}$  amount from the buyer's deposit and distribute it among all clients, except the buyer, and (viii) to return  $(S_{min} - |S_\cap|) \cdot \ddot{v}$  amount back to the buyer.
- (f) if one of the extractors cheated, there is a traitor, and the traitor delivered an incorrect result (in step 10), then instructs the ledger (i) to pay the honest extractor the same way it does in step 13(b)iiA, (ii) to refund the traitor  $\ddot{c}h$  amount, and (iii) to pay the buyer and the rest of the clients in the same way it does in step 16b.
- (g) outputs whatever  $\mathcal{A}$  outputs and then halts.

Next, we show that the real and ideal models are computationally indistinguishable. We first focus on the adversary's output. The addresses of the smart contracts have identical distribution in both models. In the real and ideal models, the adversary sees the transcripts of ideal calls to  $f_{\text{ctr}}$  as well as the functionality outputs  $(mk, mk')$ . Due to the security of  $\text{CT}$  (as we are in the  $f_{\text{ctr}}$ -hybrid world), the transcripts of  $f_{\text{ctr}}$  in both models have identical distribution, so have the random outputs of  $f_{\text{ctr}}$ , i.e.,  $(mk, mk')$ . Also, the deposit amounts  $S_{min} \cdot \ddot{v}$  and  $\ddot{w}$  have identical distributions in both models. Due to the semantical security of the public key encryption, the ciphertext  $ct_{mk}$  in the real model is computationally indistinguishable from the ciphertext  $ct_{mk}$  in the ideal model. Due to the hiding property of the commitment scheme, commitment  $com_{mk}$  in the real model is computationally indistinguishable from commitment  $com_{mk}$  in the ideal model. Moreover, due to the security of  $\text{JUS}$ , all transcripts and outputs produced in the ideal model, in step 7 above, have identical distribution to the corresponding transcripts and outputs produced in  $\text{JUS}$  in the real model. The address of  $\mathcal{SC}_{\text{TC}}$  has the same distribution in both models. The amounts each party receives in the real and ideal models are the same, except when both extractors produce an identical and incorrect result (i.e., intersection) in the real model, as we will shortly discuss, this would not occur under the assumption that the extractors are rational and due to the security of the counter collusion smart contracts.

Now, we show that an honest party aborts with the same probability in the real and ideal models. As before, for the sake of completeness, we include the  $\text{JUS}$  in the following discussion as well. Due to the security of  $\text{CT}$ , an honest party, during  $\text{CT}$  invocation, aborts with the same probability in both models; in this case, the adversary learns nothing about the parties' input set and the sets' intersection as the parties have not sent out any encoded input set yet. In both models, an honest party can read the smart contract and check if sufficient amounts of coins have been deposited. Thus, it would halt with the same probability in both models. If the parties halt because of insufficient amounts

of deposit, no one could learn about (i) the parties' input set and (ii) the sets' intersection because the inputs (representation) have not been dispatched at this point. Due to the security of ZSPA-A, an honest party during ZSPA-A execution aborts with the same probability in both models. In this case, an aborting adversary also learns nothing about the parties' input set and the sets' intersection.

Due to the security of VOPR, honest parties abort with the same probability in both models. In the case where a party aborts during the execution of VOPR, the adversary would learn nothing (i) about its counter party's input set, and (ii) about the rest of the honest parties' input sets and the intersection as the other parties' input sets remain blinded by random blinding factors known only to client  $D$ . In the real model, client  $D$  can check if all parties provided their encoded inputs via reading the state of the smart contract. The simulator can perform the same check to ensure  $\mathcal{A}'$  has provided the encoded inputs of all corrupt parties. So, in both models, an honest party with the same probability detects if not all encoded inputs have been provided. In this case, if an adversary aborts and does not provide its encoded inputs (i.e., polynomials  $\nu^{(C'')}$ ), then it learns nothing about the honest parties' input sets and the intersection, for the same reason explained above.

In the real model, the contract sums every client  $C$ 's polynomial  $\nu^{(C)}$  with each other and with client  $D$ 's polynomial  $\nu^{(D)}$ , that ultimately removes the blinding factors that  $D$  initially inserted (during the VOPR execution), and then checks if the result is divisible by  $\zeta$ . Due to (a) Theorem 5, (b) the fact that the smart contract is given the random polynomial  $\zeta$  in plaintext, (c) no party (except honest  $D$ ) knew polynomial  $\zeta$  before they send their input to the contract, and (d) the security of the contract (i.e., the adversary cannot influence the correctness of the smart contract's verifications), the contract can detect if a set of outputs of VOPR were tampered with, with a probability at least  $1 - \epsilon(\lambda)$ . In the ideal model,  $\text{Sim}_A^{\text{ANE}}$  (in step 7o) can remove the blinding factors and it knows the random polynomial  $\zeta$ . So,  $\text{Sim}_A^{\text{ANE}}$  can detect when  $\mathcal{A}'$  tampers with a set of the outputs of VOPR (sent to  $\text{Sim}_A^{\text{ANE}}$ ) with a probability at least  $1 - \epsilon(\lambda)$ , due to Theorem 5. Therefore, the smart contract in the real model and the simulator in the ideal model would abort with a similar probability.

Due to the security of ZSPA-A, the probability that in the real model an invalid  $k_i \in \vec{k}$  is appended to  $L$  is similar to the probability that  $\text{Sim}_A^{\text{ANE}}$  detects an invalid  $k'_i \in \vec{k}'$  in the ideal model. In the real model, when  $\text{Flag} = \text{False}$ , the smart contract can identify each ill-structured output of VOPR (i.e.,  $\nu^{(C)}$ ) with a probability of at least  $1 - \epsilon(\lambda)$  by checking whether  $\zeta$  divides  $\nu^{(C)}$ , due to (a) Theorem 4 (i.e., unforgeable polynomial), (b) the fact that the smart contract is given  $\zeta$  in plaintext, (c) no party (except honest client  $D$ ) knew anything about  $\zeta$  before they send their input to the contract, and (d) the security of the contract. In the ideal model, when  $\text{Flag} = \text{False}$ , given each  $\nu^{(C'')}$ ,  $\text{Sim}_A^{\text{ANE}}$  can remove its blinding factors from  $\nu^{(C'')}$  which results in  $\phi^{(C'')}$  and then can check if  $\zeta$  divides  $\phi^{(C'')}$ , in step 7(q)vi.  $\text{Sim}_A^{\text{ANE}}$  can detect an ill-structured  $\nu^{(C'')}$  with a probability of at least  $1 - \epsilon(\lambda)$ , due to Theorem 4, the fact that the simulator is given  $\zeta$  in plaintext, and the adversary is not given any knowledge about  $\zeta$  before

it sends to the simulator the outputs of **VOPR**. Therefore, the smart contract in the real model and  $\text{Sim}_A^{\text{ANE}}$  in the ideal model can detect an ill-structured input of an adversary with the same probability. The smart contract in the real model and  $\text{Sim}_A^{\text{ANE}}$  in the ideal model can detect and abort with the same probability if the adversary provides an invalid opening to each commitment  $\text{com}_{i,j}$  and  $\text{com}_{mk}$ , due to the binding property of the commitment scheme. Also, the smart contract in the real model and  $\text{Sim}_A^{\text{ANE}}$  in the ideal model, can abort with the same probability if a Merkle tree proof is invalid, due to the security of the Merkle tree, i.e., due to the collision resistance of Merkle tree's hash function.

Note that in the ideal model,  $\text{Sim}_A^{\text{ANE}}$  can detect and abort with a probability of 1, if  $\mathcal{A}_r$  does not send to the simulator all encoded encrypted elements of the intersection, i.e., when  $E^{(I)} \neq S'$ . Because the simulator already knows all elements in the intersection (and the encryption key). Thus, it can detect with a probability of 1 if both the intersection sets that the extractors provide are identical but incorrect. In the real world, if the extractors collude with each other and provide identical but incorrect intersections, then an honest client (or the smart contract) cannot detect it. Thus, the adversary can distinguish the two models, based on the probability of aborting. However, under the assumption that the smart contracts (of Dong *et al.* [17]) are secure (i.e., are counter-collusion), and the extractors are rational, such an event (i.e., providing identical but incorrect result without one extractor betraying the other) would not occur in either model, as the real model and  $(\mathcal{A}_1, \mathcal{A}_2)$  rational adversaries follow the strategy that leads to a higher payoff. Specifically, as shown in [17], providing incorrect but identical results is not the preferred strategy of the extractors; instead, the betrayal of one extractor by the other is the most profitable strategy in the case of (enforceable) collusion between the two extractors. This also implies that the amounts that the extractors would receive in both models are identical.

Now, we analyse the output of the predicates  $\bar{Q} := (Q^{\text{Init}}, Q_R^{\text{Del}}, Q_R^{\text{UF-A}}, Q^{\text{F-A}})$  in the real and ideal models. In the real model, all clients proceed to prepare their input set only if the predefined amount of coins have been deposited by the parties; otherwise (if in steps 2,4,5 of **ANE** and step 4 of **JUS** there is not enough deposit), they will be refunded and the protocol halts. In the ideal model, the simulator proceeds to prepare its inputs only if enough deposit has been placed in the contract. Otherwise, it would send message  $\text{abort}_1$  to TTP, during steps 2–7c. Thus, in both models, the parties proceed to prepare their inputs only if  $Q^{\text{Init}}(.) \rightarrow 1$ . In the real model, if there is an abort after the parties ensure there is enough deposit and before client  $D$  provides its encoded input to the contract, then all parties can retrieve their deposit in full. In this case, the aborting adversary cannot learn anything about honest parties' input sets, as the parties' input sets have been blinded by random blinding polynomials known only to client  $D$ . In the ideal model, if there is any abort during steps 7e–7k, then the simulator sends  $\text{abort}_2$  to TTP and instructs the ledger to refund every party's deposit. In the case of an abort, within the above two steps, the auditor is not involved, and paid. Therefore, in both models, in the case of an abort within the above steps, we would have  $Q^{\text{F-A}}(.) \rightarrow 1$ .



In the real model, if  $Flag = True$ , then all parties can locally extract the intersection, regardless of the extractors' behaviour. In this case, each honest party receives  $\ddot{y}'$  amount that it initially deposited in  $\mathcal{SC}_{JUS}$ . Moreover, each honest party receives *at least*  $|S_{\cap}| \cdot \ddot{l}$  amount as a reward, for contributing to the result. In this case, the honest buyer always collects the leftover of its deposit. Specifically, if both extractors act honestly, and the intersection cardinality is smaller than  $|S_{min}|$ , then the buyer collects its deposit leftover, after paying all honest parties. If any extractor misbehaves, then the honest buyer fully recovers its deposit (and the misbehaving extractor pays the rest). Even in the case that an extractor misbehaves and then becomes a traitor to correct its past misbehaviour, the buyer collects its deposit leftover if the intersection cardinality is smaller than  $|S_{min}|$ . In the ideal model, when  $Flag = True$ , then  $\text{Sim}_A^{\text{ANE}}$  can extract the intersection by summing the output of VOPR provided by all parties and removing the blinding polynomials. In this case, it sends back each party's deposit placed in  $\mathcal{SC}_{JUS}$ , i.e.,  $\ddot{y}'$  amount. Also, in this case, each honest party receives at least  $|S_{\cap}| \cdot \ddot{l}$  amount as a reward and the honest buyer always collects the leftover of its deposit. Thus, in both models in the case of  $Flag = True$ , we would have  $Q_R^{\text{Del}}(.) \rightarrow 1$ .

In the real model, when  $Flag = False$ , only the adversary can learn the result. In this case, the contract sends (i)  $\ddot{c}h$  amount to  $Aud$ , and (ii)  $\frac{m' \cdot \ddot{y}' - \ddot{c}h}{m - m'}$  amount, as compensation and reward, to each honest party, in addition to each party's initial deposit. In the ideal model, when  $Flag = False$ ,  $\text{Sim}_A^{\text{ANE}}$  sends  $abort_3$  to TTP and instructs the ledger to distribute the same amount the contract distributes among the auditor (e.g., with address  $adr_j$ ) and every honest party (e.g., with address  $adr_i$ ) in the real model. Thus, in both models when  $Flag = False$ , we would have  $Q_R^{\text{UF-A}}(., ., ., ., adr_i) \rightarrow (a = 1, .)$  and  $Q_R^{\text{UF-A}}(., ., ., ., adr_j) \rightarrow (., b = 1)$ .

We conclude that the distribution of the joint outputs of the honest client  $C \in \hat{G}$ , client  $D$ ,  $Aud$ , and the adversary in the real and ideal models are computationally indistinguishable.

**Case 2: Corrupt dealer  $D$ .** In the real execution, the dealer's view is defined as follows:

$$\begin{aligned} \text{View}_D^{\text{ANE}}(S^{(D)}, (S^{(1)}, \dots, S^{(m)})) = \\ \{S^{(D)}, adr_{sc}, S_{min} \cdot \ddot{v}, 2 \cdot \ddot{d}', r_D, \text{View}_D^{\text{JUS}}, \text{View}_D^{\text{CT}}, (com_{1,j}, \bar{e}_1, q_1, h_1) \dots, \\ (com_{sz,j'}, \bar{e}_{sz}, q_{sz}, h_{sz}), g, \hat{x} := (mk, z'), S_{\cap}\} \end{aligned}$$

where  $\text{View}_D^{\text{CT}}$  and  $\text{View}_D^{\text{VOPR}}$  refer to  $D$ 's real-model view during the execution of CT and VOPR respectively. Also,  $r_D$  is the outcome of internal random coins of  $D$ ,  $adr_{sc}$  is the address of contract,  $\mathcal{SC}_{ANE}$ ,  $(j, \dots, j') \in \{1, \dots, h\}$ ,  $z' = \text{PRF}(\bar{m}k, 0)$ ,  $sz = |S_{\cap}|$ , and  $h_i$  is a Merkle tree proof asserting that  $com_{i,j}$  is a leaf node of a Merkle tree with root node  $g$ . The simulator  $\text{Sim}_D^{\text{ANE}}$ , which receives all parties' input sets, works as follows.

1. generates an empty view. It appends to the view, the input set  $S^{(D)}$ . It constructs and deploys a smart contract. It appends the contract's address,  $adr_{sc}$ , to the view.
2. appends to the view integer  $S_{min} \cdot \ddot{v}$ , and  $2 \cdot \ddot{d}'$ . Also, it appends uniformly random coins  $r'_D$  to the view.
3. extracts the simulation of JUS from JUS's simulator for client  $D$ . Let  $\text{Sim}_D^{\text{JUS}}$  be the simulation, that also includes a random key  $mk$ . It appends  $\text{Sim}_D^{\text{JUS}}$  to the view.
4. extracts the simulation of CT from CT's simulator, yielding the simulation  $\text{Sim}_D^{\text{CT}}$  that includes its output  $mk'$ . It appends  $\text{Sim}_D^{\text{CT}}$  to the view.
5. encrypts each element  $s_{i,j}$  in the intersection set  $S_\cap$  as  $e_{i,j} = \text{PRP}(mk', s_{i,j})$  and then encodes the result as  $\bar{e}_{i,j} = e_{i,j} || \text{H}(e_{i,j})$ . It commits to each encoded value as  $com_{i,j} = \text{Com}(\bar{e}_{i,j}, q_{i,j})$ , where  $j$  is the index of the bin to which  $\bar{e}_{i,j}$  belongs and  $q_{i,j}$  is a random value.
6. constructs  $(com'_{1,1}, \dots, com'_{d,h})$  where each  $com'_{i,j}$  is a value picked uniformly at random from the commitment scheme output range. For every  $j$ -th bin, it sets each  $com'_{i',j}$  to unique  $com_{i,j}$  if value  $com_{i,j}$  for  $j$ -th bin has been generated in step 5. Otherwise, the original values of  $com'_{i',j}$  remains unchanged.
7. constructs a Merkle tree on top of the values  $(com'_{1,1}, \dots, com'_{d,h})$  generated in step 6. Let  $g$  be the resulting tree's root.
8. for each element  $s_{i,j}$  in the intersection, it constructs  $(com'_{i',j}, \bar{e}_{i,j}, q_{i,j}, h_{i,j})$ , where  $com'_{i',j}$  is the commitment of  $\bar{e}_{i,j}$ ,  $com'_{i',j} \in com$ ,  $(\bar{e}_{i,j}, q_{i,j})$  is the commitment's opening generated in step 5, and  $h_{i,j}$  is a Merkle tree proof asserting that  $com'_{i',j}$  is a leaf of a Merkle tree with root  $g$ . It appends all  $(com'_{i',j}, \bar{e}_{i,j}, q_{i,j}, h_{i,j})$  along with  $g$  to the view.
9. generates a commitment to  $mk$  as  $com_{mk} = \text{Com}(mk, z')$  where  $z' = \text{PRF}(mk, 0)$ . It appends  $(mk, z')$  and  $S_\cap$  to the view.

Now, we will discuss why the two views are computationally indistinguishable.  $D$ 's input  $S^{(D)}$  is identical in both models, so they have identical distributions. The contract's address has the same distribution in both models. The same holds for the integers  $S_{min} \cdot \ddot{v}$  and  $2 \cdot \ddot{d}'$ . Also, because the real-model semi-honest adversary samples its randomness according to the protocol's description, the random coins in both models (i.e.,  $r_D$  and  $r'_D$ ) have identical distribution. Due to the security of JUS,  $\text{View}_D^{\text{JUS}}$  and  $\text{Sim}_D^{\text{JUS}}$  have identical distribution, so do  $\text{View}_D^{\text{CT}}$  and  $\text{Sim}_D^{\text{CT}}$  due to the security of CT. The intersection elements in both models have identical distributions and the encryption scheme is schematically secure. Therefore, each intersection element's representation (i.e.,  $\bar{e}_i$  in the real model and  $\bar{e}_{i,j}$  in the deal model) are computationally indistinguishable. Each  $q_i$  in the real model and  $q_{i,j}$  in the ideal model have identical distributions as they have been picked uniformly at random. Each commitment  $com_{i,j}$  in the real model is computationally indistinguishable from commitment  $com'_{i',j}$  in the ideal model.

In the real model, each Merkle tree proof  $h_i$  contains two leaf nodes (along with intermediary nodes that are the hash values of a subset of leaf nodes) that are themselves the commitment values. Also, for each  $h_i$ , only one of the leaf node's openings (that contains an element in the intersection) is seen by  $D$ . The

same holds in the ideal model, with the difference that for each Merkle tree proof  $h_{i,j}$  the leaf node whose opening is not provided is a random value, instead of an actual commitment. However, due to the hiding property of the commitment scheme, in the real and ideal models, these two leaf nodes (whose openings are not provided) and accordingly the two proofs are computationally indistinguishable. In both models, values  $mk$  and  $z'$  are random values, so they have identical distributions. Furthermore, the intersection  $S_\cap$  is identical in both models. Thus, we conclude that the two views are computationally indistinguishable.

**Case 3: Corrupt auditor.** In this case, the real-model view of the auditor is defined as

$$\begin{aligned} \text{View}_{Aud}^{\text{ANE}} \left( \perp, S^{(D)}, (S^{(1)}, \dots, S^{(m)}) \right) = \\ \{ \text{View}_{Aud}^{\text{JUS}}, \text{adr}_{sc}, S_{min} \cdot \ddot{v}, 2 \cdot \ddot{d}', (com_{1,j}, \bar{e}_1, q_1, h_1) \dots, (com_{sz,j'}, \bar{e}_{sz}, q_{sz}, \\ h_{sz}), g, \hat{x} := (mk, z') \} \end{aligned}$$

Due to the security of JUS,  $Aud$ 's view  $\text{View}_{Aud}^{\text{JUS}}$  during the execution of JUS can be easily simulated. As we have shown in Case 2, for the remaining transcript, its real-model view can be simulated too. However, there is a difference between Case 3 and Case 2; namely, in the former case,  $Aud$  does not have the PRP's key  $mk'$  used to encrypt each set element. However, due to the security of PRP, it cannot distinguish each encrypted encoded element from a uniformly random element and cannot distinguish  $\text{PRP}(mk', \cdot)$  from a uniform permutation. Therefore, each value  $\bar{e}_j$  in the real model has identical distribution to each value  $\bar{e}_{i,j}$  (as defined in Case 2) in the ideal model, as both are the outputs of PRP.

**Case 4: Corrupt public.** In the real model, the view of the public (i.e., non-participants of the protocol) is defined as below:

$$\begin{aligned} \text{View}_{Pub}^{\text{ANE}} \left( \perp, S^{(D)}, (S^{(A_1)}, \dots, S^{(A_m)}) \right) = \\ \{ \text{View}_{Pub}^{\text{JUS}}, \text{adr}_{sc}, S_{min} \cdot \ddot{v}, 2 \cdot \ddot{d}', (com_{1,j}, \bar{e}_1, q_1, h_1) \dots, (com_{sz,j'}, \bar{e}_{sz}, q_{sz}, \\ h_{sz}), g, \hat{x} := (mk, z') \} \end{aligned}$$

Due to the security of JUS, the public's view  $\text{View}_{Pub}^{\text{JUS}}$  during JUS's execution can be simulated in the same way which is done in Case 4, in Section O. The rest of the public's view overlaps with  $Aud$ 's view in Case 3, excluding  $\text{View}_{Aud}^{\text{JUS}}$ . Therefore, we can use the argument provided in Case 3 to show that the rest of the public's view can be simulated. We conclude that the public's real and ideal views are computationally indistinguishable.

## T Further Discussion on Anesidora

### T.1 Strawman Approach

There is a simpler but costlier approach to finding the intersection without involving the extractors; that is the smart contract finds the (encoded) elements of the intersection and distributes the parties' deposit according to the number of elements it finds. This approach is simpler, as we do not need the involvement of (i) the extractors and (ii) the three counter collusion contracts. Nevertheless, it is costlier, because the contract itself needs to factorise the unblinded resulting polynomial and find the roots, which would cost it  $O(d^2)$  for each bin, where  $d$  is the size of each bin. Our proposed approach however moves such a computation off-chain, leading to a lower monetary computation cost.

### T.2 Encrypting Set Elements

In ANE, each party initially encrypts its set elements (using deterministic encryption) to preserve the privacy of the elements from non-participants of the protocol, e.g., the public. This allows extractors to prove to the (public) smart contract  $\mathcal{SC}_{\text{ANE}}$  that they have already committed to the encryption of the elements in the intersection without revealing the actual plaintext elements.

### T.3 The Use of the Hash-based Padding

The reason each client uses the hash-based padding to encode each encrypted element  $e_i$  as  $\bar{e}_i = e_i || \mathbf{H}(e_i)$  is to allow the auditor in the counter collusion contracts to find the error-free intersection, without having to access to one of the original (encrypted) sets.

### T.4 Extracting the Elements in the Intersection

Compared to JUS, there is a minor difference in finding the result in ANE. Specifically, because in ANE each set element  $s_i$  is encoded as (i)  $e_i = \text{PRP}(mk', s_i)$  and then (ii)  $\bar{e}_i = e_i || \mathbf{H}(e_i)$  by a client, then when the client wants to find the intersection it needs to first regenerate  $\bar{e}_i$  as above and then treat it as a set element to check if  $\phi'(\bar{e}_i) = 0$ , in step 12(b)iii of JUS.

### T.5 The Use of Double-layered Commitments

In ANE, each extractor uses double-layered commitments (i.e., it first commits to the encryption of each element and then constructs a Merkle tree on top of all commitments) for efficiency and privacy purposes. Constructing a Merkle tree on top of the commitments allows the extractor to store only a single value in  $\mathcal{SC}_{\text{ANE}}$  would impose a much lower storage cost compared to the case where it would store all commitments in  $\mathcal{SC}_{\text{ANE}}$ . Also, committing to the elements' encryption lets it hide from other clients the encryption of those elements that are not in

the intersection. Recall that encrypting each element is not sufficient to protect one client's elements from the rest of the clients, as they all know the decryption key.

## T.6 Inserting Garbage Inputs

To increase their reward, malicious clients may be tempted to insert “garbage” elements into their sets with the hope that those garbage elements appear in the result and accordingly they receive a higher reward. However, they would not succeed as long as there exists a semi-honest client (e.g., dealer  $D$ ) which uses actual set elements. In this case, by the set intersection definition, those garbage elements will not appear in the intersection.

## T.7 Using Generic Reward Function

In ANE, for the sake of simplicity, we let each party receive a fixed reward, i.e.,  $\tilde{l}$ , for every element it contributes to the intersection. It is possible to make the process more flexible/generic. For instance, we could define a Reward Function  $RF$  that takes  $\tilde{l}$ , an (encoded) set element  $e_i$  in the intersection, its distribution/value  $val_{e_i}$ , and output a reward  $rew_{e_i}$  that each party should receive for contributing that element to the intersection, i.e.,  $RF(\tilde{l}, e_i, val_{e_i}) \rightarrow rew_{e_i}$ .

## U Proof of Theorem 9

Below, we restate the proof of Theorem 9, taken from [4].

*Proof.* Let  $P = \{p_1, \dots, p_t\}$  and  $Q = \{q_1, \dots, q_{t'}\}$  be the roots of polynomials  $\mathbf{p}$  and  $\mathbf{q}$  respectively. By the Polynomial Remainder Theorem, polynomials  $\mathbf{p}$  and  $\mathbf{q}$  can be written as  $\mathbf{p}(x) = \mathbf{g}(x) \cdot \prod_{i=1}^t (x - p_i)$  and  $\mathbf{q}(x) = \mathbf{g}'(x) \cdot \prod_{i=1}^{t'} (x - q_i)$  respectively, where  $\mathbf{g}(x)$  has degree  $d - t$  and  $\mathbf{g}'(x)$  has degree  $d' - t'$ . Let the product of the two polynomials be  $\mathbf{r}(x) = \mathbf{p}(x) \cdot \mathbf{q}(x)$ . For every  $p_i \in P$ , it holds that  $\mathbf{r}(p_i) = 0$ . Because (a) there exists no non-constant polynomial in  $\mathbb{F}_p[X]$  that has a multiplicative inverse (so it could cancel out factor  $(x - p_i)$  of  $\mathbf{p}(x)$ ) and (b)  $p_i$  is a root of  $\mathbf{p}(x)$ . The same argument can be used to show for every  $q_i \in Q$ , it holds that  $\mathbf{r}(q_i) = 0$ . Thus,  $\mathbf{r}(x)$  preserves roots of both  $\mathbf{p}$  and  $\mathbf{q}$ . Moreover,  $\mathbf{r}$  does not have any other roots (than  $P$  and  $Q$ ). In particular, if  $\mathbf{r}(\alpha) = 0$ , then  $\mathbf{p}(\alpha) \cdot \mathbf{q}(\alpha) = 0$ . Since there is no non-trivial divisors of zero in  $\mathbb{F}_p[X]$  (as it is an integral domain), it must hold that either  $\mathbf{p}(\alpha) = 0$  or  $\mathbf{q}(\alpha) = 0$ . Hence,  $\alpha \in P$  or  $\alpha \in Q$ .