AVeCQ: Anonymous Verifiable Crowdsourcing with Worker Qualities

Sankarshan Damle
IIIT Hyderabad
sankarshan.damle@research.iiit.ac.in

Vlasis Koutsos
HKUST
vkoutsos@cse.ust.hk

Dimitrios Papadopoulos
HKUST
dipapado@cse.ust.hk

Dimitris Chatzopoulos
Univeristy College of Dublin
dimitris.chatzopoulos@ucd.ie

Sujit Gujar
IIIT Hyderabad
sujit.gujar@iiit.ac.in

ABSTRACT

In crowdsourcing systems, requesters publish tasks and interested workers provide answers to get rewards. *Worker anonymity* motivates participation since it protects their privacy. *Anonymity with unlinkability* is an enhanced version of anonymity because it makes it impossible to "link" workers across the tasks they participate in. Another core feature of crowdsourcing systems is *worker quality* which expresses a worker's trustworthiness and quantifies their historical performance. Notably, worker quality depends on the participation history, revealing information about it, while unlinkability aims to disassociate the workers' identities from their past activity. In this work, we present AVeCQ, the first crowdsourcing system that reconciles these properties, achieving enhanced anonymity and verifiable worker quality updates. AVeCQ relies on a suite of cryptographic tools, such as zero-knowledge proofs, to *(i)* guarantee workers' privacy, *(ii)* prove the correctness of worker quality scores and task answers, and *(iii)* commensurate payments. AVeCQ is developed in a modular way where the requesters and workers communicate over a platform that supports pseudonymity, information logging, and payments. In order to compare AVeCQ with the state-of-the-art we prototyped it over Ethereum. AVeCQ outperforms the state-of-the-art in three popular crowdsourcing tasks (image annotation, average review, and Gallup polls). For instance, for an Average Review task with 5 choices and 128 participating workers AVeCQ is 40% faster (including overhead to compute and verify the necessary proofs, and blockchain transaction processing time) with the task's requester consuming 87% fewer gas units.

PVLDB Reference Format:
Sankarshan Damle, Vlasis Koutsos, Dimitrios Papadopoulos, Dimitris Chatzopoulos, and Sujit Gujar. AVeCQ: Anonymous Verifiable Crowdsourcing with Worker Qualities. PVLDB, 14(1): XXX-XXX, 2020.
doi:XX.XX/XXX.XX

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at github.com/sankarshandamle/AVeCQ.

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.doi:XX.XX/XXX.XX

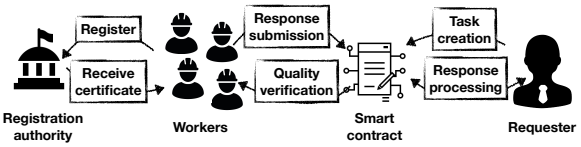

Figure 1: The crowdsourcing setting and phases of AVeCQ

1 INTRODUCTION

Crowdsourcing is the process of gathering information regarding a *task* (e.g., a query or project) by leveraging a set of agents who are incentivized to work on them within a specific time frame [29]. A prominent example of crowdsourcing revolves around Human Intelligence Tasks (HITs), which can be used to enrich datasets designed for empowering machine learning models. Those who request crowdsourcing tasks can extract statistical data, form conclusions, and even monetize from any results based on the individually-provided answers.

Specifically, a popular use case is the calculation of the average over a set of values. Representative examples can be found in personal data analytics (e.g., average salary calculation), smart agriculture (average crop collection), smart grid (average daily energy consumption), and others [26, 62, 75, 82]. Other motivating tasks revolve around calculating a set's n-most popular items. E.g., for $n = 1$ this encompasses image annotation [39, 60], while for $n > 1$ Gallup polls [53].

More concretely, a *requester* publishes a task seeking information and *workers* provide their responses. The requester can define a *task policy* specifying various task parameters (e.g., the task description, a final answer calculation mechanism, a minimum number of participating workers). To *incentivize* participation, requesters may compensate workers [19, 50, 69]. E.g., the workers may be compensated based on a flat rate or even based on how "close" their responses are to the final answer of the task. Such compensation mechanisms can be specified in the task policy and in fact, there exists a plethora of deployed crowdsourcing systems that operate in this paradigm (e.g., Amazon MTurk [1], Microwork [4], and QMarkets [5]).

A common feature in crowdsourcing systems is that workers are individually associated with a *quality score*, which estimates how "trustworthy" their responses are based on prior performance [52]. Indeed, worker qualities can be a vital tool for a requester, who can use them to screen workers (e.g., specifying a certain threshold so that only a worker whose quality surpasses it may participate), or even pay them according to their quality scores. A worker's

quality score (we refer to it as *quality*) is dynamic, as it may change after every participation, depending on the task policy. In practice, qualities represent the performance of workers in previous tasks. Thus, they can ultimately assist in "ostracizing" workers who submit answers without judiciously performing tasks [49, 64, 69]. E.g., workers who submit arbitrary answers will have their quality decrease over time, making it increasingly harder to clear quality thresholds and participate in future tasks.

Privacy in Crowdsourcing. Protecting worker's private information is greatly important [34]. In practice, task participation may require workers to disclose sensitive data to requesters (e.g., location, age, gender, race). Being able to observe the answer pattern of a specific worker is therefore undesirable; in fact, it opens the system up to worker profiling [34] and potential discrimination! For example, suppose Alice deploys a task requesting workers to disclose their racial background. After Bob provides such information, solely based on this, Alice may choose to exclude Bob from her future tasks.

Motivated by this, a line of works has emerged that studies privacy in crowdsourcing systems and proposes corresponding solutions [20, 31, 41, 45, 57, 66, 70, 72, 73]. The required property in these works is *worker anonymity*: it should be impossible to deduce a worker's identity from information revealed *while participating in a task*. A "naive" way to achieve anonymity would be to hide workers' identities behind pseudonyms. However, this is not enough as, through a series of tasks, requesters may still be able to identify workers on the basis of their answers alone. If a worker participates in multiple tasks, requesters may be able to build a "rich" profile *linked* to a certain pseudonym. To avoid such a case, prior works [42, 45] consider a stronger privacy notion, *anonymity with unlinkability*: It should be impossible to link a worker's participation across tasks. Throughout the rest of the paper we refer to anonymity as in this "stronger" variant.

We now make the following observations based on the discussion above, regarding qualities and anonymity. On one hand, qualities directly stem from workers' participation profiles. On the other hand, anonymity aims to obscure all past participation information. Thus, the two properties appear to be *inherently contradictory*: achieving one seemingly precludes the other. In fact, there exist works that achieve anonymous crowdsourcing without worker quality [45, 57, 66] and vice versa [41] (see Table 1). To the best of our knowledge, no prior work simultaneously achieves both.

This Work. We propose AVeCQ (Figure 1), a crowdsourcing system that is *the first to satisfy worker anonymity while maintaining worker qualities in a verifiable manner*. Table 1 highlights the differences between AVeCQ and existing works in terms of achieved properties (see Section 2 for a more in-depth comparison).

First, our system achieves anonymity with unlinkability, i.e., requesters learn *nothing* about participating workers except for their explicit task answers and possibly that their corresponding qualities are above the threshold specified in the task's policy (but, crucially, not the quality itself). Second, even though each worker's participation history remains hidden, AVeCQ supports *verifiable* qualities, meaning that the following two conditions apply. To participate in a task, workers must prove they are using their *correctly calculated* quality as *derived* by their entire participation history. Likewise,

	Worker Quality	Anonymity	Policy Verifiable	Data Confidentiality	Sybil Resistance	Free-rider Resistance
Yan *et al.* [78]	○	○	●†	●	●	●
Duan *et al.* [17]	○	○	●‡	●	○	●
PACE [81]	○★	○	●†	●	○	●†
TrustWorker [22]	○★	○	●†	○	●	○
Dragoon [46]	○★	○	●	●	●	●
bHIT [43]	○★	○	●	●	●	●
zkCrowd [83]	○	●	●‡/♦	●	●	●
ZebraLancer [45]	○	●	●	●	●	●
BPRF [32]	●	◐	●♦	○	●	●
AVeCQ	●	●	●	●	●	●

○/●/◐: Absence/Presence/Weaker-version of the property,
★: Data quality, †: Semi-honest intermediary, ‡: TEE, ♦: Blockchain

Table 1: AVeCQ vs related works. Special symbols denote weaker variants of the property are achieved.

upon task completion, requesters must prove the correctness of the participants' quality updates according to the answers and the task policy. Crucially, the workers in AVeCQ verify their updated qualities without any information about the final answer except for what can be trivially inferred from the policy. Note that achieving each of this properties on its own is rather straight-forward (e.g., if we do not care about protecting worker identities it is easy to check that the correct quality score is used for each task); the challenge arises when simultaneously trying to achieve both.

AVeCQ is additionally secure against other significant threats. Anonymity might allow workers to generate multiple identities arbitrarily (i.e., perform a Sybil attack) and reap extra payments [74]. AVeCQ avoids Sybil attacks by requiring workers certificates to be issued by a *Registration Authority* (RA) before task participation. Moreover, we counter *free-rider attacks*, i.e., workers cannot submit someone else's response or use another's quality as their own to successfully participate in a task. Last, AVeCQ guarantees fair worker compensation according to task policy, i.e., the requester cannot "cheat" to avoid payments.

Overview of Challenges and Techniques. At the core of our solution lies a wide range of cryptographic techniques, including zero-knowledge proofs (ZKPs) that allow a prover to convince a verifier about the correctness of a computation, without revealing any private information. Below we state briefly the major challenges we encountered in this work and how we overcame them.

Privately updating qualities, verifiably. To hide worker qualities from the task requester we have workers submit their qualities as additively homomorphic commitments. This enables requesters to increment/decrement all workers' qualities without ever accessing any raw underlying quality. This calculation is based on the workers' answers, the final answer, and the task policy. Now, all workers know their own answer to the task and the policy, so when they see their updated quality it is trivial to check whether the change is computed correctly or not, *if they also know the final answer*. However, since AVeCQ does not reveal the final answer to the workers, the question of how to verify that the corresponding updates are honestly computed arises. To this end, the requester is obligated to present individual ZKPs to all workers, regarding the correctness of their quality update.

Proof of quality-freshness. If a worker's quality decreases after participating in a task she may be incentivised to discard the latest quality update and reuse her earlier quality in future tasks. Thus, we face the following problem: "*How can we ensure that the quality used is always the latest, according to all previous task participations?*". To address this, we borrow and adapt a technique used in privacy-preserving cryptocurrencies [28, 37]. The proof that requesters compute for each worker after task completion must also pertain to the fact that the updated quality commitment of a worker has been appended as a leaf to a Merkle Tree which contains all quality commitments across all tasks. A worker that wishes to participate in a task provides a ZKP that pertains to the fact that the re-randomized quality commitment she provides corresponds to the quality committed in one of the Merkle tree leaves, without revealing to which one. In this manner, a worker that tries to benefit by reusing an earlier quality will break her anonymity due to the way this proof is crafted in AVeCQ—specifically as the same Merkle leaf will need to be used again (see Section 5). Crucially, this is also easy to detect even long after the worker's participation.

Anonymous task participation. To anonymously participate in future tasks, the worker cannot utilize the quality commitment in the way it exists in the Merkle tree as that would trivially break anonymity. To surpass this obstacle, a worker can re-randomize her quality commitment and provide the requester with a ZKP that corresponds to its correctly updated quality and is above the task participation threshold. However, a new question arises now: "*how can a requester be certain that the worker has not re-randomized an outdated quality commitment?*" To prevent such behavior, each worker must submit a cryptographic hash, including the quality commitment used in the ZKP above, concatenated with the unique identifier it provided to the RA during registration. Thus, any worker trying to re-randomize/use a previous quality could be trivially detected.

Nevertheless, if the hash is computed "naively" now the RA can launch possible de-anonymization attacks. To avoid this, essentially, the worker must not hash any public information (i.e., its latest quality commitment transmitted by the latest task requester) together with its identifier. Instead, the requester, upon concluding its task, will transmit the updated quality of the worker combined with a "dummy" commitment, whose randmoness will communicate to the worker in a confidential manner. The worker can then decompose the commitment, extract its *true* quality commitment, and continue participating in tasks without compromising its identity.

Resistance to free-riding. Workers might attempt to participate effortlessly by (re)submitting someone else's response or quality. AVeCQ resists such free-riding attacks by tying each worker's answer to her quality and a unique "payment wallet" via a ZKP. Thus, no worker attempting to "hijack" someone else's response is able to produce valid proof for participation and get compensated.

Implementing AVeCQ. We implement a prototype of AVeCQ , whose smart contract component is deployed over Ethereum testnets, i.e., Rinkeby [6] and Goerli [3]. To demonstrate the practicality and scalability of AVeCQ, we report extensively on its performance focusing on computational and blockchain overheads, communication bandwidth, and monetary costs. We test AVeCQ on three popular, real-world-inspired tasks [39, 62, 82], i.e., image annotation, average review estimation, and Gallup polls, using the real datasets Duck [75], Amazon Review [26], and COVID-19 Survey [53]. Our results show, somewhat surprisingly, that AVeCQ outperforms other state-of-the-art systems, even though it achieves a combination of stronger security properties and/or operates in a stronger security model (Table 1). We benchmark AVeCQ's performance for representative tasks against the "best" systems with available implementation. For instance, for binary image annotation with 39 workers our end-to-end (E2E) time is < 8 minutes, vs. < 2 hours for only 11 workers in [45]. Note that contrary to ZebraLancer, AVeCQ's smart contract is solely used for storage purposes (e.g., no on-chain verfication happens—see Section 5). AVeCQ also retains its edge in terms of gas consumption. E.g., for generating an average review with 128 workers, a requester in AVeCQ consumes $\approx$ 4.3M gas units, whereas in [17] the requester requires $\approx$ 35M gas units for just 100 workers. Last, AVeCQ consumes $< 25\%$ of the gas required in [46] (See Section 7 for a more detailed comparison).

Our Contributions. In summary, we construct a crowdsourcing system that bridges the gap between anonymity and worker qualities while being able to scale to real-world inspired task instances. We highlight the main contributions of our work as follows:

(1) We design AVeCQ, the first crowdsourcing system that guarantees the anonymity of participating workers across tasks while maintaining a quality system verifiably (see Section 5). Crucially, it does so without compromising functionality, as it can support arbitrary policies and tasks.
(2) We provide definitions for three critical security and privacy properties of crowdsourcing systems: anonymity, free-rider resistance, and policy verifiability. We prove that AVeCQ satisfies all three properties under standard assumptions (see Section 6)[1].
(3) We develop a prototype implementation of AVeCQ and test its performance thoroughly. In terms of efficiency and scalability, our implementation is comparable, when not better, than other state-of-the-art systems providing less functionality (e.g., support only gold-standard tasks) or operate in a weaker threat model (see Section 7). In fact, we provide an in-depth comparison with prior works, both qualitative (in terms of properties) and quantitative (in terms of three specific real-world tasks).

2 RELATED WORK

Worker Quality. Prior works in crowdsourcing systems adopt different notions of "quality". The authors of [22, 43, 46, 81] interpret worker quality in terms of proximity to an "estimated" or "final" answer. Specifically, Lu et al. [46] use a set of gold-standard tasks, whose final answer is known a priori to the requester and a posteriori to the workers, to determine the quality of the answers. Despite being commonly used, this approach is rather limiting since it only works when the answer is known. It does not work for other popular crowdsourcing tasks e.g., Gallup polls. The authors of [22, 81] assign scores to workers based on the proximity of their response to the mean of the submitted data. Contrary, (as also in [32, 41, 68, 69]) we interpret worker quality as a representation of workers' entire historical task performance. This is not only more realistic but also strictly more general, since AVeCQ can capture all previous quality notions through different task policies.

[1] Additionally, in Appendix A, we show AVeCQ to be secure against other, various, popular attacks.

Privacy. Another point of contention in the literature revolves around the definition of identity and data privacy—both of which are essential. Anonymity with unlinkability protects the workers identities across tasks entirely, as explained before. Contrary, data privacy limits what the system entities (e.g., blockchain nodes) learn about the workers' data.

Anonymity requires two conditions to be met: *(i)* the identity of a worker remains hidden and *(ii)* the quality itself cannot be used to de-anonymize workers. The authors of [24, 42] satisfy *(i)* by enabling workers to generate identities freely. However, this in turn allows workers to carry out Sybil attacks. Similarly, CrowdBC [41] suffers from the same limitation, with two additional drawbacks. Crucially, CrowdBC stores qualities on-chain, in the plain. This poses a potential and significant risk to de-anonymizing workers, as explained before, rendering [41] not anonymous.

Regarding *data privacy* towards third parties (most commonly referred to as *confidentiality*), it is commonplace to require workers to submit their responses in an encrypted manner as done in [9, 17, 30, 41, 44–46, 73, 81], with two exceptions. First, in [22] a deterministic encryption scheme is used meaning the Computing Server (the intermediary collecting worker responses) can access them in the plain, while in [32] all responses are already communicated in the plain. The authors of [20, 31, 47, 65–67, 70, 71] use differential privacy to protect workers' inputs. However, noisy methods affect the correctness of the crowdsourcing process as they dilute the final answer. Thus, they are impractical when including qualities in the crowdsourcing model, since most quality update mechanisms are based on correlations between a worker's answer and the "final answer" of a task, and even more so when the set of possible answers is of limited size.

Verifiable Policy. To achieve policy verifiability, works such as [17, 22, 32, 78, 81, 83] either entrust *(i)* a semi-honest intermediary, *(ii)* a Trusted Execution Environment (TEE), e.g., Intel SGX, or *(iii)* blockchain miners to carry out policy-related computations (e.g., calculate the final answer or rewards). However, having to "blindly" trust intermediaries is not ideal, secure hardware is susceptible to side-channel attacks [10, 11], and on-chain computations jeopardise data confidentiality. AVeCQ is free of any such assumptions and is policy-verifiable under only cryptographic assumptions.

Other Security Issues. Two common security threats in crowdsourcing are *Sybil* and *Free-riding* attacks. A countermeasure to Sybil attacks is employing a trusted RA that registers workers by issuing a certificate based on the worker's unique identifier (e.g., ID documentation) [32, 44, 45, 63, 83]. In fact, the authors of [17, 22, 41] do not utilize an RA and fail to prevent such attacks. Alternatively, in [17] the authors argue that their *incentive-compatible* mechanism disincentives worker misbehavior —a strictly stronger assumption.

To safeguard against free-riding, requesters can employ a trusted third party [79], or couple workers' certificates with their public addresses [45]. CrowdBC [41], alternatively, requires workers to deposit funds to a smart contract to be eligible for a task. Workers are incentivized to exert effort, or risk their deposits. Instead, AVeCQ eliminates free-riding attacks also by relying only on cryptographic assumptions.

AVeCQ **vs. State-of-the-art.** Unlike the works above, our system satisfies all properties in Table 1. Recently, Liang et al. proposed bHIT [43], a blockchain-based crowdsourcing system for HITs that, similarly to AVeCQ, does not disclose the responses of the workers to the public blockchain. However, bHIT does not provide any notion of anonymity or policy verifiability and operates in a weaker security model since it does not consider colluding workers.

Zebralancer [45] is the closest work to ours in terms of the employed techniques and properties. Both AVeCQ and Zebralancer utilize zk-SNARKS but Zebralancer only uses them for proving the correctness of the rewards calculation, while in AVeCQ zk-SNARKS are also used by workers to prove that they are using their latest valid qualities. Moreover, requesters use zk-SNARKS to prove the correctness of the final answer, updates of qualities, and calculation of payments. Similarly, both works employ an RA, and utilize smart contracts for task deployment and participation. However, in contrast to Zebralancer, AVeCQ additionally supports worker qualities. This extra feature is far from trivial to implement as new security and privacy concerns emerge resulting in a more complex protocol. Nevertheless, this does not come at a performance cost, as AVeCQ is at least as (or even more) efficient than Zebralancer, despite of supporting worker qualities.

3 PRELIMINARIES

We now present the tools used in AVeCQ. Table 2 provides a reference for key notation. Let $\mathbb{E}$ be an elliptic curve defined over a large prime field $\mathbb{F}_p$ with $G, H \in \mathbb{E}$ as publicly known generators. We denote by $x \leftarrow\!\$\ A$ the sampling at random of the element x from the domain A. We denote by λ a security parameter and by $negl(\lambda)$ a function negligible in λ. Last, we denote by $Adv^{\mathcal{G}}(\mathcal{A})$ the advantage that $\mathcal{A}$ has in winning the $\mathcal{G}$ game.

Pedersen Commitments [51]. A commitment scheme binds and hides a value x. Specifically, a Pedersen commitment of x with randomness r is in the form of $\mathsf{Com}(x, r) = x \cdot G + r \cdot H$. Pedersen commitments are *additively homomorphic*, i.e., $\mathsf{Com}(x_1, r_1) + \mathsf{Com}(x_2, r_2) = \mathsf{Com}(x_1 + x_2, r_1 + r_2)$, *computationally binding* (it is not feasible to "change one's mind" after committing), and *perfectly hiding* (they reveal nothing about the committed data).

Public-Key Encryption (PKE) Scheme [36]. A PKE scheme consists of the following algorithms:

- $\mathsf{KeyGen}(\lambda) \rightarrow (sk, pk)$. Given the security parameter λ, KeyGen samples a secret key $sk \leftarrow\!\$\ \{0,1\}^{\lambda}$, computes the public key $pk = sk \cdot G$. It outputs the key-pair (sk, pk).
- $\mathsf{Enc}(pk, x; r) \rightarrow E(pk, x; r)$. To encrypt a value x, the algorithm takes input a randomness r and outputs the curve point $(r \cdot G, P_x + \cdot r(sk \cdot G))$. Here, P_x is a publicly-known mapping of a value x to a curve point in $\mathbb{E}$.
- $\mathsf{Dec}(E(pk, x; r), sk) \rightarrow x$. To decrypt x from $E(pk, x; r)$, the algorithms computes $x := P_x + \cdot r(sk \cdot G) - r \cdot sk \cdot G$.

Hash Function [15]. A cryptographic hash function $H : \{0,1\}^* \rightarrow \{0,1\}^{\lambda}$ is collision-resistant if the probability of two distinct inputs mapping to the same output is negligible: $\Pr[H(x) = H(y) \mid x \neq y] \leq negl(\lambda)$. Additionally, it is pre-image resistant if the probability of inverting it is negligible. We denote the indistinguishability games for these properties as $H - CR$ and $H - PR$ respectively.

Crowdsourcing Model	
$\mathcal{A}^\tau$	Set of possible answers for task τ
a^τ_ϕ, P^τ	Final answer and task policy for τ
d^τ, $\mathcal{DL}^\tau$	Description and deadlines of τ
q^τ_i	Quality score of w_i at τ
n^τ_{th}	Threshold number of workers required for τ
n^τ	Total number of worker responses for τ
Protocol Notations	
$cert_i$	EdDSA signature for party i
$root_{MT}$	Root of a Merkle Tree (MT)
$path_{leaf}$	path for $leaf$ in MT
$r^\tau_{k,i}$	w_i's randomness for τ
$r^\tau_{*,i}$	w_i's randomness to re-randomize $\mathrm{Com}(q^{\tau-1}_i, \cdot)$
$r^\tau_{**,i}$	Randomness of the dummy commitment for q^τ_i
$r^\tau_{c,i}$	Randomness with $\mathrm{Com}(0, r^\tau_{**,i})$ for q^τ_i
$r^\tau_{c-d,i}$	Randomness without $\mathrm{Com}(0, r^\tau_{**,i})$ for q^τ_i
$E(pk_R, a^\tau_i; \cdot)$	Encryption of w_i's response a^τ_i for τ
$E(pk_R, pa^\tau_i; \cdot)$	Encryption of w_i's address pa^τ_i for τ
$\mathrm{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i} + r_{\star,i})$	Re-randomized commitment of q_i for τ
$E(pk_R, r^\tau_{k,i}; \cdot)$	Encryption of w_i's index $r^\tau_{k,i}$ for τ
$H\left(\mathrm{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i}), m_i\right)$	Quality "tag" of w_i for τ
zk-SNARKs	
$\pi^\tau_{o_i}$	ProveQual proof for w_i's tuple o^τ_i for τ
$\pi^\tau_{a_\phi}$	AuthCalc proof for a^τ_ϕ for τ
$\pi^\tau_{q_i}$	AuthQual proof for w_i's updated q^τ_i for τ
$\pi^\tau_{a_i}$	AuthValue proof for w_i's response a^τ_i for τ

Table 2: Key Notations

Digital Signatures [8]. A digital signature scheme allows verification of the authenticity of a certificate. EdDSA is a Schnorr-based signature scheme defined over $\mathbb{E}$. In EdDSA, given G, one derives its public key pk by sampling $sk \leftarrow\$ \mathbb{F}_p$. A party i signs the value $H(m_i)$ for a secret message m_i denoted as its signature $cert_i = (R_i, S_i)$. Here, $R_i = r \cdot G$ s.t. $r \leftarrow\$ \mathbb{F}_p$, and $S_i = r + H(m_i) \cdot sk$. A verifier accepts the signature iff $S_i \cdot G = R_i + H(m_i) \cdot pk$ holds.

zk-SNARKs [58]. A zero-knowledge succinct non-interactive argument of knowledge (zk-SNARK) allows a prover to convince a verifier about the correctness of a computation on private input, through a protocol. The verifier-available information is referred to as *statement* $\vec{x}$ and the private input of the prover as *witness* $\vec{\omega}$. The protocol execution takes place in a non-interactive manner, with succinct communication. A zk-SNARK consists of: *(i)* a *Setup* algorithm which outputs the public parameters PP for a NP-complete language $\mathcal{L}_\mathcal{S} = \{\vec{x} \mid \exists\ \vec{\omega}$ s.t. $\mathcal{S}(\vec{x}, \vec{\omega}) = 1\}$, where $\mathcal{S} : \mathbb{F}^n \times \mathbb{F}^h \rightarrow \mathbb{F}^l$ is the *arithmetic circuit satisfiability problem* of an $\mathbb{F}$-arithmetic circuit; *(ii)* A *Prover* algorithm that outputs a constant size proof π, attesting to the correctness of $\vec{x} \in \mathcal{L}_\mathcal{S}$ with witness $\vec{\omega}$; and *(iii)* A *Verifier* algorithm which efficiently checks the proof. Informally, a zk-SNARK satisfies the following properties:

- *Completeness.* If $\exists \vec{\omega} : \mathcal{L}_\mathcal{S}(\vec{x}, \vec{\omega}) = 1$, honest Provers can always convince Verifiers.
- *Soundness.* If $\exists \vec{\omega}' : \mathcal{L}_\mathcal{S}(\vec{x}, \vec{\omega}) = 0$, a dishonest Prover has negligible probability in convincing the Verifier.
- *Zero-knowledge.* If $\exists \vec{\omega} : \mathcal{L}_\mathcal{S}(\vec{x}, \vec{\omega}) = 1$, the Verifier does not learn any information about $\vec{\omega}$ (besides its existence).

Merkle Tree (MT) [48]. A Merkle tree (MT) is a complete binary tree where each parent node is a hash of its children. This structure allows for *membership* proofs, attesting to the existence of a specific leaf via a publicly known root ($root_{MT}$) and a path ($path_{leaf}$).

Blockchain & Smart Contracts. A blockchain is a ledger distributed across peers and made secure through cryptography and incentives. Peers agree upon storing information in the form of blocks through consensus algorithms. Blockchain technology has transcended its use in cryptocurrency applications, especially with the introduction of smart contracts with Ethereum [77]. A smart contract is a computer program that can be run in an on-chain manner. Performing any smart-contract computation over Ethereum requires *gas*. The amount of gas charged depends on the type of computation. More computationally extensive operations require higher gas to be executed on-chain. The total charge for the computation is referred to as *gas cost*. Any computation that alters the state of the contract, i.e., alters any contract method or variables, consumes gas. Contrary, reading data from the contract is free. To submit state-altering transactions users specify a *gas price* that they are willing to pay per *gas unit*.

EIP-1559 [13]. In Ethereum, each transaction creator pays a dynamic base fee b and a priority fee δ (in gas). Each block's miner receives δ while the base fee is "burned" (i.e., removed from the supply, forever). δ affects the verification time, as higher δ results in faster transaction verification time.

Crowdsourcing Quality-Update Policies. Various techniques have been proposed for updating workers' qualities after task participation [12, 14, 35, 61, 80]. AVeCQ integrates the widely adopted in academia [14, 80] and industry (Amazon MTurk [1]) *Beta* distribution. Formally, for each w_i and task τ, we maintain integers α^τ_i and β^τ_i, initialised to 1. The quality score for τ is then given by the Beta distribution with mean $\frac{\alpha^\tau_i}{\alpha^\tau_i + \beta^\tau_i}$. The update rule considering a worker's response a^τ_i and the "final answer" a^τ_ϕ is:

- if $a^\tau_i = a^\tau_\phi : \alpha^{\tau+1}_i = \alpha^\tau_i + 1;\ \beta^{\tau+1}_i = \beta^\tau_i$
- if $a^\tau_i \neq a^\tau_\phi : \beta^{\tau+1} = \beta^\tau + 1;\ \alpha^{\tau+1}_i = \alpha^\tau_i$

AVeCQ uses this mechanism to handle quality updates and specifically computes increments/decrements as Pedersen commitments of 1 and 0, accordingly. That way a worker can utilize the commitments' additive homomorphic property to compute its new quality.

4 PROBLEM FORMULATION

In this section, we present the problem formulation in three aspects: system model, threat model, and problem statement.

System Model. It includes three types of entities, namely a *Registration Authority* (RA), *a set of Requesters*, and *a set of Workers*. The RA is in charge of registering workers in the system by providing them with a participation certificate, upon receiving a unique identifier. Requesters can *(i)* create and publish tasks, and *(ii)* collect worker-generated responses. Workers observe published tasks and upon wishing to participate in a task they provide the respective requester with their individual responses. At a high level, to create a task, a requester needs to disclose its description alongside an answer-calculation mechanism, a quality-update rule, and a payment scheme. Remember that each worker has a *quality* that reflects their trustworthiness based on their previous answers. Specifically, we denote all registered workers by the set $\mathcal{W} = \{w_1, \ldots, w_n\}$, and their associated qualities by the set $\mathcal{Q} = \{q_1, \ldots, q_n\}$.

Entity	Adversarial Behavior
Requester	– "de-anonymize" workers from multi-task participation – update qualities and calculate payments arbitrarily – avoid paying or updating qualities
Worker	– participate in a single task multiple times – use other worker's answer/quality and get rewards
RA	– track a worker across tasks

Table 3: Possible attacks based on AVeCQ 's threat model

Threat Model. We make no assumptions as to the behavior of requesters or workers and we allow the entities to collude arbitrarily. Malicious requesters (R) can try to infer information about participating workers (not trivially leaked by their answers), calculate the final answer and the updated qualities arbitrarily, and avoid payments. On the other hand, malicious workers (W) can try to generate multiple identities arbitrarily and respond by utilizing outdated quality scores or even someone else's response. Last, the RA is considered to be semi-honest, but may try to track a worker across tasks. Based on the above, there exists a plethora of attacks to systems operating in our threat model (see Table 3).

Problem Statement. Considering the above system and threat model, the problem we study in this paper is the following: How to design an efficient and scalable system that carries out arbitrary crowdsourcing tasks and supports worker qualities, while safeguarding against the mentioned attacks.

5 AVeCQ

In this section we present our crowdsourcing system. Our construction utilizes cryptographic components (i.e. digital signature, public-key encryption, commitment scheme, zero-knowledge proof protocol, Merkle tree, and hash function) and a smart contract. Particularly, to create a task τ, a requester can specify a set of attributes depending on the expressivity of the task. These include: *(i)* the task description d^τ, *(ii)* the set of available answer choices $\mathcal{A}^\tau = \{a_1^\tau, \ldots, a_c^\tau\}$, *(iii)* the maximum budget Γ^τ the requester is willing to allocate to the workers, *(iv)* a set of deadlines $\mathcal{DL}^\tau$ (signifying until when workers may submit their responses and until when task processing must be completed by the requester), *(v)* a threshold number of workers n_{th}^τ (denoting the minimum participation of workers that is needed to calculate the final answer), *(vi)* the requester's public key pk_R (which the workers will use to encrypt their sensitive data), and *(vii)* the task policy P^τ (including task-related information e.g., quality-threshold participation requirements, the final answer calculation mechanism, the quality-update rule, and the payment scheme). We denote the set of responses workers submit to task τ as $O_{n^\tau}^\tau = \{o_1^\tau, \ldots, o_{n^\tau}^\tau\}$, where n^τ denotes the the total workers who provided responses to τ. A response o_i of worker w_i includes, among others, its answer to the task (a_i^τ) and its quality from the previous task ($q_i^{\tau-1}$)[2]. Therefore, we denote the set of answers to the task τ as $A_{n^\tau}^\tau = \{a_1^\tau, \ldots, a_{n^\tau}^\tau\}$ and the set of corresponding qualities as $Q_{n^\tau}^{\tau-1} = \{q_1^{\tau-1}, \ldots, q_{n^\tau}^{\tau-1}\}$.

During a task, the requester collects $O_{n^\tau}^\tau$ and extracts $A_{n^\tau}^\tau$. Then, (if needed) it calculates the final answer and for all participants their updated qualities and payments. For this, the requester uses the following algorithms:

(1) $\text{AnsCalc}(A_{n^\tau}^\tau, Q_{n^\tau}^{\tau-1}, P^\tau) \rightarrow a_\phi^\tau$: On input the set of answers from the participating workers $A_{n^\tau}^\tau$, the qualities $Q_{n^\tau}^{\tau-1}$ and the participating policy P^τ, it outputs the final answer a_ϕ^τ.

(2) $\text{QualCalc}(Q_{n^\tau}^{\tau-1}, a_\phi^\tau, P^\tau) \rightarrow Q_{n^\tau}^\tau$: On input the set of the quality scores from the participants $Q_{n^\tau}^{\tau-1}$, the final answer a_ϕ^τ, and the participation policy P^τ, outputs the set of the updated quality scores for every participating worker $Q_{n^\tau}^\tau = \{q_1^\tau, \ldots, q_{n^\tau}^\tau\}$.

(3) $\text{PaymCalc}(A_{n^\tau}^\tau, Q_{n^\tau}^{\tau-1}, a_\phi^\tau, P^\tau) \rightarrow Paym^\tau$: On input the set of the answers $A_{n^\tau}^\tau$ and qualities $Q_{n^\tau}^{\tau-1}$, the final answer a_ϕ^τ, and the participation policy P^τ, outputs the set of payments for every worker that participated $Paym^\tau = \{p_1^\tau, \ldots, p_n^\tau\}$.

5.1 Protocol

Next we describe our construction and explain the design rationale. Particularly, our protocol comprises of two stages: a preprocessing-setup and a task-specific one. To assist the reader, we provide the following task as a representative example and follow its execution through the rest of this subsection.

Running example (τ_{ex}). A requester deploys on 31/1/2023 at 23 : 00 the following image annotation task:

$$\mathcal{AT} = \{\text{"Does this image contain a duck?"}, (\text{Yes}, \text{No}),$$
$$200ETH, (31/1/2023/23:30, 1/2/2023/23:59), 1001,$$
$$\texttt{0x7584e47e7a7e09a6be64dc3aeaf0b64364234d9c},$$
$$(\text{Quality} > 75\%, \text{Majority}, \text{Beta distribution},$$
$$\text{Correct:0.0001ETH,Not correct:0.00005ETH})\}$$

Remark: This is the most expressive task which can be executed through AVeCQ, in terms of supported features. For tasks requiring less functionality (e.g. reward each worker horizontally or do not require/need qualities) the related elements/steps can be omitted.

Preprocessing-Setup stage. Ahead of time, the RA generates the public parameters PP by running the zk-SNARKs' setup and encryption key-generation algorithms. Now, when a worker w_i wishes to participate in the crowdsourcing system, it presents the RA with unique identification data m_i. The RA, in turn, generates a participation certificate $cert_i$ (i.e., an EdDSA signature on m_i, which in the U.S.A. could be the Social Security Number of w_i.). Additionally, it initializes the quality of w_i, generates the commitment $\text{Com}(q_i^{(0)}, r_{c,i}^{(0)})$ where $r_{c,i}^{(0)} \leftarrow\!\$\ \mathbb{Z}_p$, and appends this as a leaf to a Merkle Tree MT. Last, the RA provides w_i with $(cert_i, q_i^{(0)}, r_{c,i}^{(0)})$. Notably, this is a *dynamic* stage, as new workers can register arbitrarily and independently of other operations. In τ_{ex}, the RA sets up the SNARKs for calculating the most popular answer a_ϕ^τ, the Beta distribution, and for the $a_i^\tau \stackrel{?}{=} a_\phi^\tau$ check. We provide more details on all four zk-SNARKs construction AVeCQ utilizes in Section 5.2, with further details in Appendix B.

Task-specific stage. To carry out a specific task, the requester and participating workers engage in a protocol having *four* phases: *Task Creation, Response Submission, Response Processing, and Quality Verification*. Broadly, a requester creates a crowdsourcing task

[2] We acknowledge that the "previous" task for two workers w_i and w_j might differ, however we use $q_i^{\tau-1}$ and $q_j^{\tau-1}$ to denote their latest qualities.

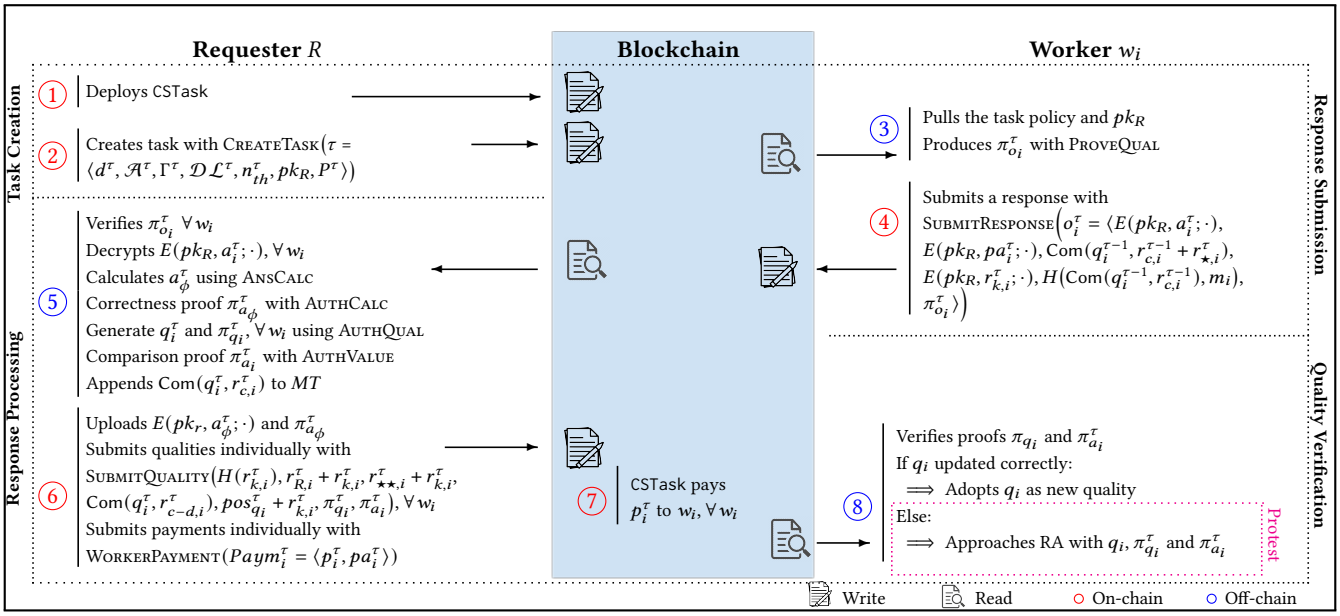

Figure 2: AVecQ: Task-specific stage analysis. To initiate a task, a requester R deploys a smart contract. A worker may submit a response including her (encrypted) answer, latest quality (commitment), and attest to the validity of the quality and conformity to the task policy (via a ZKP). R then verifies all submitted proofs off-chain, computes the final answer a_ϕ^τ, quality updates (commitments), payments, and ZKPs, and uploads all but a_ϕ^τ on-chain. Last, workers get rewards and adopt their new qualities.

by deploying a smart contract on a blockchain to elicit responses from interested workers. Upon collecting all responses, the requester invokes AnsCalc to extract the final answer. Additionally, the requester invokes QualCalc to calculate the updated worker qualities and PaymCalc to calculate individual compensations. All participants have known public addresses and can connect to the blockchain network. Figure 2 depicts all task-specific phases of AVeCQ's protocol. We present all the phases in detail, below.

5.1.1 Task Creation. This phase includes solely on-chain computations. First, a requester deploys a smart contract (CSTask) and deposits Γ^τ funds into it. Next, to publish a new task τ the requester (suppose Alice) uses the CreateTask method of CSTask and uploads on-chain the following transaction: $tx_{\text{CreateTask}(\tau)} = \langle d^\tau, \mathcal{A}^\tau, \Gamma^\tau, \mathcal{DL}^\tau, n_{th}^\tau, pk_R, P^\tau\rangle = \mathcal{AT}$, in τ_{ex}. The requester then awaits for the next phase to conclude.

5.1.2 Response Submission. An interested w_i can invoke the SubmitResponse method of CSTask to submit its response. Naively, w_i just needs to provide the requester with its answer, its quality, a proof about holding a valid quality, and a public address for compensation. However, SubmitResponse-related data are uploaded to a blockchain. Thus, if we allow w_i to send this data in the plain some of our security properties will be trivially violated[3].

To ensure no leakage of sensitive data, w_i provides a response o_i^τ containing, *(i)* an encryption of its answer a_i^τ, *(ii)* an encryption of its public address pa_i^τ, *(iii)* a re-randomized commitment regarding its latest quality $q_i^{\tau-1}$, *(iv)* an encryption of a random value $r_{k,i}^\tau$, *(v)* a hash of the commitment in MT with m_i, and *(vi)* a corresponding proof of correctness for all these values. This phase includes off-chain and on-chain computations as w_i generates commitments, encryptions, hashes, and proofs locally; and later on invokes the SubmitResponse method to upload o_i^τ on CSTask.

Off-chain. First, w_i calculates the ciphertexts for its answer, address, and a random value $r_{k,i}^\tau \leftarrow\$ \mathbb{Z}_p$: $E(pk_R, a_i^\tau;\cdot)$, $E(pk_R, pa_i^\tau;\cdot)$, $E(pk_R, r_{k,i}^\tau;\cdot)$[4]. Additionally, w_i computes the commitment Com$(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^\tau)$[5] and the hash $H\left(\text{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\right)$[6] where $r_{\star,i}^\tau \leftarrow\$ \mathbb{Z}_p$. Now, using the proof generating algorithm of our ProveQual zk-SNARK, a worker w_i produces a proof $\pi_{o_i}^\tau$. Specifically, $\pi_{o_i}^\tau$ attests to *(i)* w_i having registered, *(ii)* the existence of a leaf in MT that hides $q_i^{\tau-1}$, *(iii)* $q_i^{\tau-1}$ conforms to P^τ, and *(iv)* $(m_i, q_i^{\tau-1})$ are included in the calculation of $H\left(\text{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\right)$.

On-chain. A worker w_i can use the SubmitResponse method of CSTask to submit a response o_i^τ in the form of the following transaction: $tx_{SubmitResponse}(o_i^\tau) = \langle E(pk_R, a_i^\tau;\cdot), E(pk_R, pa_i^\tau;\cdot)$, Com$(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^\tau), E(pk_R, r_{k,i}^\tau;\cdot), H\left(\text{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\right), \pi_{o_i^\tau}\rangle$.

5.1.3 Response Processing. During this phase, the requester computes a_ϕ^τ, computes and communicates (via the SubmitQuality method) the updated worker qualities, and initiates payments (via WorkerPayment). Notably, workers need to be certain that the requester updated qualities, and corresponding payments, based on P^τ. Thus, requesters provide individual zk-SNARK proofs for correctly updating qualities. Overall, the requester performs the following off-chain and on-chain computations.

[3] See Appendix A.5 for a more elaborate analysis

[4] Crucially, $r_{k,i}^\tau$ will be used to hide the leaf-position of the worker's newly updated quality commitment. [5] Com$(q_i^{\tau-1}, r_{c,i}^{\tau-1})$ is a leaf in MT. [6] We utilize this hash as a *tag* that ensures that w_i cannot re-submit a quality without being detected.

Off-chain. After the response submission deadline (specified in $\mathcal{DL}^\tau$) has passed and $n^\tau \geq n^\tau_{th}$ workers have submitted responses, the requester uses the verification algorithm of ProveQual to verify all proofs $\pi_{o_i^\tau}$, individually. Then, it calculates $a^\tau_\phi \leftarrow$ AnsCalc($Q^{\tau-1}_{n^\tau}, A^\tau_{n^\tau}, P^\tau$), all updated qualities using QualCalc($Q^{\tau-1}_{n^\tau}$, a^τ_ϕ, P^τ), and payments using PaymCalc($Q^{\tau-1}_{n^\tau}, A^\tau_{n^\tau}, a^\tau_\phi, P^\tau$). Recall that quality updates must be verifiable, even when only the requester knows a^τ_ϕ. To achieve this we follow the next three steps.

First, the requester generates a proof of the final answer calculation, using the workers' on-chain responses and the task policy. Specifically, it uses the zk-SNARK AuthCalc to do so. AuthCalc decrypts all encrypted answers and using P^τ calculates a^τ_ϕ. Second, the requester provides an individual "proof of correctness" for each worker w_i regarding the correlation between a^τ_ϕ, a^τ_i, and P^τ, using AuthValue. This proof includes a single decryption and comparison. Third, the requester uses o^τ_i, a^τ_ϕ, and P^τ to generate an individual "proof of correct quality update" for each worker via AuthQual. This includes a decryption and comparison as before, and additionally the verification of the new quality, based on P^τ.

Quality Updates. Recall that workers submit their qualities in the form of commitments. To update the quality of w_i, the requester computes: $\mathsf{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i} + r^\tau_{\star,i}) \circ \mathsf{Com}(\mu, r^\tau_{R,i})$, $r^\tau_{R,i} \leftarrow\!\!\$\ \mathbb{Z}_p$, where $\mu \in \mathbb{Z}$ is the difference between the old and new quality of w_i ($q^\tau_i = q^{\tau-1}_i + \mu$). Finally, for each w_i, the requester re-randomizes the newly generated qualities commitment using "dummy" commitments i.e., $\forall i \in [n^\tau]\mathsf{Com}(0, r^\tau_{\star\star,i}), r^\tau_{\star\star,i} \leftarrow\!\!\$\ \mathbb{Z}_p$. Formally, the requester appends the commitment $\mathsf{Com}(q^\tau_i, r^\tau_{c,i}) = \mathsf{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i} + r^\tau_{\star,i}) \circ \mathsf{Com}(\mu, r^\tau_{R,i}) \circ \mathsf{Com}(0, r^\tau_{\star\star,i})$ to the MT and we denote $r^\tau_{c,i} = r^{\tau-1}_{c,i} + r^\tau_{\star,i} + r^\tau_{R,i} + r^\tau_{\star\star,i}$. We also denote $\mathsf{Com}(q^\tau_i, r^\tau_{c-d,i}) = \mathsf{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i} + r^\tau_{\star,i}) \circ \mathsf{Com}(\mu, r^\tau_{R,i})$ as the quality commitment without the "dummy" commitment with $r^\tau_{c-d,i} = r^{\tau-1}_{c,i} + r^\tau_{\star,i} + r^\tau_{R,i}$. In τ_{ex}, suppose Bob is a worker who has answered "Yes", along with the majority of the rest of the workers, meaning a^τ_ϕ="Yes". Alice computes $\mathsf{Com}(q^\tau_{i,\alpha}, r^\tau_{c,i}) = \mathsf{Com}(q^{\tau-1}_{i,\alpha}, \cdot) \times \mathsf{Com}(1, \cdot) \times \mathsf{Com}(0, \cdot)$ and $\mathsf{Com}(q^\tau_{i,\beta}, r^\tau_{c,i}) = \mathsf{Com}(q^{\tau-1}_{i,\beta}, \cdot) \times \mathsf{Com}(1, \cdot) \times \mathsf{Com}(0, \cdot)$.

On-chain. The requester invokes the SubmitQuality method, to communicate the newly updated qualities to each worker w_i. The transaction is of the form: $tx_{SubmitQuality} = \langle H(r^\tau_{k,i}), r^\tau_{R,i} + r^\tau_{k,i}, r^\tau_{\star\star,i} + r^\tau_{k,i}, \mathsf{Com}(q^\tau_i, r^\tau_{c-d,i})$[7], $pos^\tau_{q_i} + r^\tau_{k,i}, \pi^\tau_{q_i}, \pi^\tau_{a_i} \rangle$. It includes: *(i)* a worker index $H(r^\tau_{k,i})$, *(ii)* the new quality randomness $r^\tau_{R,i} + r^\tau_{k,i}$, *(iii)* the randomness of the "dummy" commitment $r^\tau_{\star\star,i} + r^\tau_{k,i}$, *(iv)* the commitment of the updated quality $\mathsf{Com}(q^\tau_i, r^\tau_{c,i})$, *(v)* the position $pos^\tau_{q_i} + r^\tau_{k,i}$ of the *MT* leaf storing w_i's new commitment-quality, and *(vi)* the proofs $\pi^\tau_{q_i}$, $\pi^\tau_{a_i}$ for AuthQual and AuthValue. Last, the requester submits individual payments p^τ_i through the WorkerPayment method and CSTask reimburses the workers.

5.1.4 Quality Verification. Quality verification includes only off-chain computations. In fact, w_i verifies $\pi^\tau_{q_i}$ and $\pi^\tau_{a_i}$ and adopts the updated quality. We analyze the case when proofs do not pass verification in Appendix A.5.

[7] The quality commitment without the "dummy" re-randomization.

Remark: Below-Threshold Worker Participation: Recall that a requester, upon creating a task can specify a minimum participation of n^τ_{th} workers. If $n^\tau < n^\tau_{th}$ the task is considered void and the requester compensates workers for any expenses already made, re-randomizes the quality commitments, and produces corresponding π_{q_i}; allowing workers to take part in their next task seamlessly.

5.2 zk-SNARKs

AVeCQ utilizes four different zk-SNARKs, allowing workers to participate in tasks honestly and requesters to calculate the final answer, updated qualities, and payments correctly, all of which verifiably. Below we explain the functionality of all checks included in the employed zk-SNARKS and we include all statements, witnesses, and languages in Figure 3.

ProveQual To participate in τ, worker w_i uses ProveQual's proving algorithm to generate a proof $\pi^\tau_{o_i}$. ProveQual performs seven checks. EdDSAVer checks w_i has already registered by verifying $cert_i$ signature using m_i, while MPathVer that $q^{\tau-1}_i$ exists hidden as a commitment in a leaf of MT. Additionally, TaskVer ensures $q^{\tau-1}_i$ conforms with P^τ, and HashComVer that $(m_i, q^{\tau-1}_i)$ are included in the calculation of $H\big(\mathsf{Com}(q^{\tau-1}_i, r^{\tau-1}_{c,i}), m_i\big)$. Last, ValidEnc verifies well-formedness of the ciphertexts in o^τ_i and Qualver that m_i and the same q_i are included in the hash and the commitment.

AuthCalc The requester uses AuthCalc to attest to the correctness of a^τ_ϕ. First, ValidKeyPair checks if the secret key in the witness consists a valid key-pair with the public key provided by the requester during the task's creation. Afterwards, the circuit decrypts all workers' encrypted answers and computes the final answer based on P^τ. FinAnsVer(·)= 1 if the computed final answer matches the decryption of the requester-submitted encrypted final answer.

AuthValue and AuthQual These zk-SNARKs attest to the correctness of each worker's answer and quality update, respectively. AuthValue first checks the validity of the requester's key pair with ValidKeyPair. Next, it uses EqCheck to check if the worker's encrypted response equals the final encrypted answer. Likewise, AuthQual first uses ValidKeyPair to check the key pair's well-formedness. Additionally, it uses NewQual to check if the updated qualities were computed correctly using the encrypted final answer and the worker's response. P^τ is hard-coded in the SNARK.

6 CORE SECURITY PROPERTIES

In this section, we introduce definitions, theorems, and proofs for the three core properties of our system. In Appendix A we include additional analysis on all properties AVeCQ satisfies, as well as how our system is safeguareded against various popular attacks.

Definition 1 (Anonymity with Unlinkabiltiy). *A crowdsourcing system is anonymous with unlinkability if no PPT adversary $\mathcal{A}$ has non-negligible advantage in the following game, $\mathcal{G}_{Anon}$.*

- *Initialization: $\mathcal{A}$ specifies parameters n, λ. The challenger C runs the certificate generation algorithm to register n workers such that the maximum worker set for $\mathcal{G}_{Anon}$ is $\mathcal{W}_t = \{w_1, \cdots, w_n\}$ and samples a bit $b \leftarrow\!\!\$\ \{0, 1\}$.*

ProveQual	Statement $(\vec{x}_{PQ})$: $PP, P^\tau, root_{MT}, pk_R, pk_{RA}, \mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^{\tau-1}), H\big(\mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\big), E(pk_R, a_i^\tau; \cdot), E(pk_R, pa_i^\tau; \cdot)$ Witness $(\vec{\omega}_{PQ})$: $cert_i, m_i, q_i^{\tau-1}, r_{c,i}^{\tau-1}, r_{\star,i}^\tau, r_{\star\star,i}^{\tau-1}, \mathsf{Com}(q_i^{\tau-1}, r_{c-d,i}^{\tau-1}), a_i^\tau, pa_i^\tau, path_{q_i^{\tau-1}}$ Language $\mathcal{L}_{PQ} = \Big\{ \vec{x}_{PQ} \mid \exists\ \vec{\omega}_{PQ}$ s.t. $\mathsf{EdDSAver}(pk_{RA}; cert_i, m_i) \wedge\ \mathsf{ValidEnc}\big(E(pk_R, a_i^\tau); a_i^\tau\big) \wedge \mathsf{TaskVer}(P^\tau; q_i^{\tau-1}) \wedge$ $\mathsf{QualVer}\Big(H\big(\mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\big)), \mathsf{Com}(q_i^{\tau-1}, r_{c-d,i}^{\tau-1}); Com(q_i^{\tau-1}, r_{c,i}^{\tau-1}), r_{\star\star,i}^{\tau-1}, r_{\star,i}^\tau, m_i\Big) \wedge$ $\mathsf{HashComVer}\Big(H\big(\mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\big); \mathsf{Com}(q_i^{\tau-1}, r_{c-d,i}^{\tau-1}), \mathsf{Com}(0, r_{\star\star,i}^{\tau-1}), m_i\Big) \wedge$ $\mathsf{ValidEnc}\big(E(pk_R, pa_i^\tau); pa_i^\tau\big) \wedge \mathsf{MTPathVer}\big(root_{MT}; \mathsf{Com}(0, r_{\star\star,i}^{\tau-1}), \mathsf{Com}(q_i^{\tau-1}, r_{c-d,i}^{\tau-1}), path_{q_i^{\tau-1}}, H(r_{k,i}^{\tau-1}, m_i)\big) = 1\Big\}$.
AuthCalc	Statement $(\vec{x_{AC}})$: $PP, E(pk_R, a_\phi^\tau; \cdot), \{E(pk_R, a_i^\tau; \cdot)\}_{i\in[n^\tau]}, pk_R$ Witness $(\vec{\omega}_{AC})$: sk_R Language $\mathcal{L}_{AC} = \{\vec{x}_{AC} \mid \vec{\omega}_{AC}$ s.t. $\mathsf{ValidKeyPair}(pk_R; sk_R) \wedge \mathsf{FinAnsVer}\big(P^\tau, \{E(pk_R, a_i^\tau; \cdot)\}_{i\in[n^\tau]}, E(pk_R, a_\phi^\tau; \cdot); sk_R\big) = 1\}$.
AuthQual	Statement $(\vec{x}_{AV})$: $PP, E(pk_R, a_\phi^\tau; \cdot), E(pk_R, a_i^\tau; \cdot), pk_R$ Witness $(\vec{\omega}_{AV})$: sk_R Language $\mathcal{L}_{AV} = \{\vec{x}_{AV} \mid \vec{\omega}_{AV}$ s.t. $\mathsf{EqCheck}\big(E(pk_R, a_\phi^\tau), E(pk_R, a_i^\tau), pk_R; sk_R\big) \wedge \mathsf{ValidKeyPair}(pk_R; sk_R) = 1\}$.
AuthValue	Statement $(\vec{x}_{AQ})$: $PP, E(pk_R, a_\phi^\tau; \cdot), E(pk_R, a_i^\tau; \cdot), \mathsf{Com}(q_i^\tau, r_{c-d,i}^\tau), \mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^\tau), pk_R$ Witness $(\vec{\omega}_{AV})$: sk_R Language $\mathcal{L}_{AQ} = \{\vec{x}_{AQ} \mid \vec{\omega}_{AQ}$ s.t. $\mathsf{NewQual}\big(E(pk_R, a_\phi^\tau), E(pk_R, a_i^\tau; \cdot), \mathsf{Com}(q_i^\tau, r_{c-d,i}^\tau), \mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^\tau); sk_R\big) \wedge$ $\mathsf{ValidKeyPair}(pk_R; sk_R) = 1\}$.

Figure 3: Statements, witnesses, and languages for ProveQual,AuthCalc,AuthValue, and AuthQual.

- *Corruption Queries: When $\mathcal{A}$ issues such a query, it specifies a set of workers $\mathcal{W}_c \subseteq \mathcal{W}_t$, and C $\forall w_i \in \mathcal{W}_c$ provides all respective private information to $\mathcal{A}$.*
- *Task processing: $\mathcal{A}$ specifies a task τ and a corresponding worker set $\mathcal{W}_\tau$. $C, \forall w_i \notin \mathcal{W}_c$, samples random answers and computes all necessary encryptions, commitments, and proofs using the information for all participating workers, and forwards the responses to $\mathcal{A}$. $\mathcal{A}$ computes and communicates to C all proofs regarding the participation of all $w_i \notin \mathcal{W}_c$. If C cannot verify even one of these proofs the game halts.*
- *Challenge: $\mathcal{A}$ specifies a task τ, two worker sets $\mathcal{W}_\tau, \mathcal{W}'_\tau \subseteq \mathcal{W}_t - \mathcal{W}_c$, with $|\mathcal{W}_\tau| = |\mathcal{W}'_\tau|$, and forwards $\mathcal{W}_\tau, \mathcal{W}'_\tau$ to C. If $b = 0$ then C "runs" τ using $\mathcal{W}_\tau$ or uses $\mathcal{W}'_\tau$ otherwise. Specifically, C computes all necessary encryptions, commitments, and proofs for the non-corrupted workers of the two sets and forwards corresponding responses to $\mathcal{A}$.*
- *Finalization: $\mathcal{A}$ sends $b' \in \{0, 1\}$ to C.*

$\mathcal{A}$ wins in $\mathcal{G}_{Anon}$ if $b' = b$. A naive $\mathcal{A}$, by sampling b' at random has a $\frac{1}{2}$ probability of winning. A system is anonymous if $\forall$ PPT $\mathcal{A}$, $Adv^{\mathcal{G}_{Anon}}(\mathcal{A}) = |\Pr[b' = 1|b = 1] - \Pr[b' = 1|b = 0]| \leq negl(\lambda)$.

Theorem 1. Assuming Com is computationally hiding, ProveQual computationally zero-knowledge, and H collision and preimage resistant AVeCQ is anonymous as in Def. 1.

Proof. For an adversary to non-trivially identify whether "two workers are the same" or not, one of the following conditions must be true about the adversary: *(i)* compromised the hiding property of the commitment scheme and accessed qualities in the plain, *(ii)* found a collision in the hash function, or *(iii)* compromised the zero-knowledge property of the underlying SNARK for ProveQual and accessed identity-revealing witnesses. In $\mathcal{G}_{Anon}$ the challenge query requires the Challenger to execute a task τ with either of the following two worker sets $\mathcal{W}_\tau$ and $\mathcal{W}'_\tau$, depending on the challenger bit. Our proof follows a standard hybrid argument across all possible selections of $\mathcal{W}_\tau$ and $\mathcal{W}'_\tau$, specifically over their overlap regarding workers.

We prove indistinguishability of the view of the adversary in $\mathcal{G}_{Anon}$, regardless of the challenger bit, through a series of Hybrid games over all possible $\mathcal{W}_\tau \cap \mathcal{W}'_\tau$. We denote $\mathcal{H}_j$ the hybrid where $|\mathcal{W}_\tau \cap \mathcal{W}'_\tau| = j$. Note that, when $\mathcal{W}_\tau \cap \mathcal{W}'_\tau = n$, $Adv^{\mathcal{G}_{Anon}=\mathcal{H}_n}(\mathcal{A}) = 0$. The only difference between the views of executing hybrids $\mathcal{H}_j$ and $\mathcal{H}_{j+1}$ lie in the computation of a commitment, a hash and a proof. This means that the advantage a PPT adversary $\mathcal{A}$ has in distinguishing between the two hybrids is $Adv^{\mathcal{H},j \overset{?}{\approx} j+1}(\mathcal{A}) \leq Adv^{\text{C-Hiding}}(\mathcal{A}) + Adv^{H\text{-CR}}(\mathcal{A}) + Adv^{\text{ProveQual-ZK}}(\mathcal{A})$. Since by assumption the commitment scheme is computationally hiding, the hash function is collision-resistant, and ProveQual is computationally zero-knowledge, no PPT adversary can distinguish between these two views with non-negligible advantage. To conclude the proof we apply this transformation n times from $\mathcal{H}_0$ to $\mathcal{H}_n = \mathcal{G}_{Anon}$. Since n is polynomially bound no PPT adversary has a non-negligible advantage in winning $\mathcal{G}_{Anon}$. □

Definition 2. *A crowdsourcing system is free-rider resistant if no PPT adversary $\mathcal{A}$ has non-negligible advantage in the following game $\mathcal{G}_{FRR}$.*

- *Initialization: $\mathcal{A}$ specifies parameters n, λ. The challenger C runs the certificate generation algorithm to register n workers such that the maximum worker set for $\mathcal{G}_{FRR}$ is $\mathcal{W}_t = \{w_1, \cdots, w_n\}$ and samples a bit $b \leftarrow\$ \{0, 1\}$.*
- *Corruption Queries: When $\mathcal{A}$ issues such a query, it specifies a set of workers $\mathcal{W}_c \subseteq \mathcal{W}_t$, and C $\forall w_i \in \mathcal{W}_c$ provides all respective private information to $\mathcal{A}$.*
- *Task processing: $\mathcal{A}$ defines a task τ, by specifying $\textsc{AnsCalc}^\tau$, $\textsc{QualCalc}^\tau$, and $\textsc{PaymCalc}^\tau$, with range $[minComp^\tau, maxComp^\tau]$ for the workers' compensations. $\mathcal{A}$ forwards all this information to C who $\forall w_i \notin W_c$, samples random responses, computes all necessary encryptions and proofs using the information for all participating workers, and forwards all responses $\{o_i\} \forall w_i \in \mathcal{W}_h = \mathcal{W}_t - \mathcal{W}_c$ to $\mathcal{A}$.*

- *Challenge: $\mathcal{A}$ specifies a task τ and C provides $\mathcal{A}$ with the set of responses $O = \{o_i\}$, for each $w_i \in \mathcal{W}_h$. $\mathcal{A}$ then forwards to C a response $o_j = \langle E'(pk_R, a^{\tau}_{i_1}; \cdot), E'(pk_R, pa^{\tau}_{i_2}; \cdot), Com'(q^{\tau-1}_{i_3}, r^{\tau-1}_{c,i_4} + r^{\tau}_{\star,i_5}), E'(pk_R, r^{\tau}_{k,i_6}; \cdot), H^{\star}\left(Com(q^{\tau-1}_{i_7}, r^{\tau-1}_{c,i_8}), m_{i_9}\right), pi^{\tau'}_{o_{i_{10}}}\rangle$, where $\exists i^{\dagger} \in \{i_1, i_2 \cdots, i_{10}\}$, $i^{\dagger} \in \mathcal{W}_h$, and $\tau' \neq \tau$. If $b = 0$, C outputs $p^{\tau}_i = minComp^{\tau}$ to w_i or runs $\textsc{PaymCalc}(a^{\tau}_{i_1}, \cdot) \rightarrow p^{\tau}_i$ otherwise.*
- *Finalization: $\mathcal{A}$ sends $b' \in \{0, 1\}$ to C.*

$\mathcal{A}$ wins in $\mathcal{G}_{FRR}$ if $b' = b$. A naive $\mathcal{A}$, by sampling b' at random has a $\frac{1}{2}$ probability of winning. A system is free-riding resistant if $\forall$ PPT $\mathcal{A}$, $Adv^{\mathcal{G}_{FRR}}(\mathcal{A}) = |\Pr[b' = 1|b = 1] - \Pr[b' = 1|b = 0]| \leq negl(\lambda)$.

Theorem 2. Assuming that the PKE scheme E is CPA-secure, the hash function H is collision-resistant, the commitment scheme Com is computationally hiding, and ProveQual is computationally sound, AVeCQ is free-riding resistant as in Def. 2.

Proof. To break Free-riding Resistance, an adversary has to extract information regarding w_i's answer a_i or quality q_i. Violating the first one reduces to breaking the security of the PKE scheme E. The second condition is more complex. Specifically, the adversary "wins" iff it can break the hiding property of Com, find a collision in H, or break the soundness of ProveQual. To prove indistinguishability between the view of the adversary regardless of the challenger bit we use a hybrid analysis where we change one by one the witnesses into random values while simulating proofs. The total advantage of $\mathcal{A}$ is $Adv^{\mathcal{G}_{FFR}}(\mathcal{A}) \leq Adv^{Com}(\mathcal{A}) + Adv^{CPA-security}(\mathcal{A}) + Adv^{H-collision}(\mathcal{A}) + Adv^{SNARK-soundness}(\mathcal{A}) \leq negl(\lambda)$. □

Definition 3 (Policy Verifiability). *A crowdsourcing system is policy-verifiable if both following conditions hold:*
(i) *For any task τ_i with worker set $\mathcal{W}_i = \{w_{i,1}, \cdots, w_{i,n_i}\}$, no PPT adversary can output $a^{\tau_i,\prime}_{\phi} \leftarrow \textsc{AnsCalc}(A^{\tau_i}_{n^{\tau_i}}, P^{\tau_i})$, or $\{q_{i,1}, \cdots, q_{i,n_i}\} \leftarrow \textsc{QualCalc}(Q_{n^{\tau}}, a^{\tau_i}_{\phi}, P^{\tau_i})$, and accepting proofs $\pi^{\tau_i}_{a_{\phi}}, \{\pi^{\tau_i}_{a_{i,1}}, \cdots, \pi^{\tau_i}_{a_{i,n_i}}\}$, $\{\pi^{\tau_i}_{q_{i,1}}, \cdots, \pi^{\tau_i}_{q_{i,n_i}}\}$ with more than negligible probability.*
(ii) *For any task τ_j no PPT adversary can fabricate $q^{\star}_i$, for any w_i with q_i, and an accepting proof $\pi^{\tau_j}_{o^{\star}_i} \leftarrow \textsc{ProveQual}(q^{\star}_i, \cdot)$, where $q^{\star}_i \neq q_i$ with more than negligible probability.*

Theorem 3. *Assuming AuthCalc, AuthValue, AuthQual and ProveQual are computationally zero-knowledge, Com computationally hiding and H collision-resistant,* AVeCQ *satisfies Policy Verifiability as in Def. 3.*

Proof. To break Policy Verifiability an adversary has to violate any of the two conditions. In either case, the adversary essentially needs to provide convincing results and corresponding proofs that do not satisfy the relations in Figure 3 but still pass verification. Therefore, violating the first one reduces to breaking the soundness of AuthCalc, or AuthValue, or AuthQual or the binding property of Com. The second condition is more straightforward. Specifically, the adversary "wins" iff it can break the soundness of ProveQual or can find a collision in H with non-negligible probability. That would contradict the assumptions and therefore no PPT adversary can break the policy-verifiability property of AVeCQ . □

7 EXPERIMENTAL EVALUATION

We implement a prototype of AVeCQ and report its performance. We deploy the sole on-chain component of AVeCQ, `CSTask`, over Ethereum using Solidity [16]. Additionally, we develop requester and worker Java applications that connect with the Rinkeby [6] and Goerli [3] test networks using the Web3j framework. We use the Zokrates toolbox [18] for all zk-SNARKs implementation. Last, we perform all cryptographic operations over the ALT_BN128 elliptic curve [56]. As for the answer calculation policies, we consider *(i)* Average (Avg), and *(ii)* γ-Most Frequent (γ-MF) with $\gamma \in \{1, 3\}$.

First, we measure AVeCQ's on-chain costs in gas and USD. Next, we examine the impact of varying gas prices on the verification time for th SubmitResponse method. Then, we measure the computation time and communication size of the off-chain components. Finally, we include E2E analyses that report on the total time, space, and expenses required for the completion of three real-world tasks and compare, where possible, with state-of-the-art systems.

Setup. We construct `CSTask` with bytecode size 3.8 KB with Solidity and deploy on the Rinkeby and Goerli testnets with the Web3j framework. We monitor the created transactions through Etherscan [7]. We create three different versions for AuthCalc, AuthQual, and AuthValue zk-SNARKs for our three crowdsourcing policies. For our off-chain cryptographic primitives, we use the Zokrates-accompanying pycrypto library. For hashing we use Pedersen hash [28] and ElGamal cryptosystem as the PKE scheme, which are zk-SNARK-friendly. We execute all off-chain-component experiments on a 40-core server with Intel(R) Xeon(R) CPU E5-2640 v4 @2.40GHz and 1 GB RAM per core.

7.1 On-chain Measurements

Gas Consumption and Costs. We measure the gas consumption for `CSTask` deployment and method execution. We map gas to USD, using the default gas price of 1 Gwei = 1×10^{-9} ETH, base fee 5 GWei[8] and the price of 1 ETH = 1554.89 USD (24/01/2023). Notably, deploying `CSTask` consumes $\approx$ 1.34 million gas units and costs $\approx$ 12.49 USD. Further, the method CreateTask consumes 363491 gas units costing 3.39 USD. SubmitResponse requires 394604 units and costs the participating worker 3.68 USD. Uploading AuthCalc proof consumes 120772 gas units and costs the requester 1.13 USD.

Priority Fee vs Transaction Processing Time. Transaction processing time behaves in an "inversely proportional" manner to the chosen priority fee, δ. We observe that depending on the urgency of the task, workers may opt to use gas prices other than the default. Motivated by this, we examine the average verification time (in blocks) of SubmitResponse for each $\delta \in \{0.5, 1, 1.1, 1.5, 2, 5, 10\}$ across 200 instances, and depict the result in Figure 4. As shown, for a gas price of 1.1 GWei, the average verification time on Rinkeby is 2.54 ± 0.7 blocks ($\approx$ 30 secs). With Goerli, the verification time marginally increases to 3.52 ± 0.8 blocks ($\approx$ 42) seconds. As expected, we observe that the verification time decreases as δ increases. However, the increase is minimal ($\approx$ 0.7 block across both test networks) and the standard deviation remains almost constant, for prices ranging from 1.1 GWei to 10 GWei. Importantly though, on Rinkeby,

[8] At the time of writing, the Ethereum Mainnet base fee was $\approx$ 5 GWei [2].

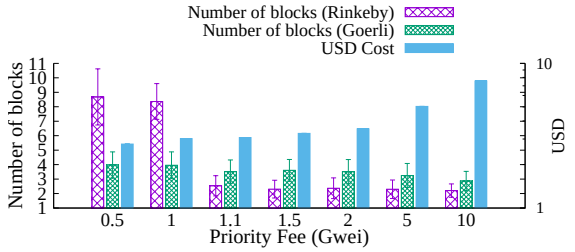

Figure 4: SubmitResponse Verification Time vs Priority Fee. Here, the base fee is 5 GWei.

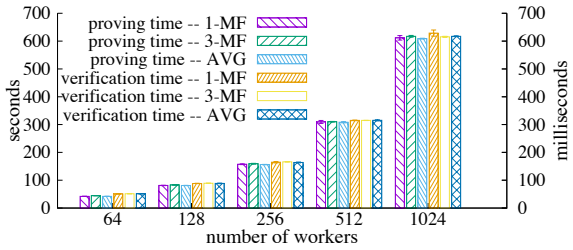

Figure 5: AuthCalc: Proving & Verification Time vs. number of workers with 4 choices.

for gas prices < 1.1 GWei, the difference in processing time is considerable and with high deviation. E.g., for 0.5 GWei the average processing time is 8.68 blocks, with a deviation of ≈ 2 blocks.

The Ethereum Main network shows a similar trend. For instance, the verification time decreases from ≈ 10 minutes with $\delta = 0$ to ≈ 3 minutes with $\delta = 1$ GWei [2]. As each crowdsourcing task's time sensitivity is absorbed in its deadline, we believe that 1 GWei is a reasonable priority fee. For shorter deadlines, a worker can accordingly increase its priority fee.

7.2 Off-chain Measurements

Non-SNARK Computations. Both the requester and the workers generate ElGamal ciphertexts, Pedersen hashes and commitments. A single encryption takes ≈ 11ms, while a decryption takes < 1ms. Constructing a pre-image and computing a Pedersen hash requires ≈ 59ms, while generating a Pedersen commitment takes < 1ms. Overall, the communication size between the requester and a worker is 384B and between a worker and the requester is 448B.

SNARKs Performance. Here we present measurements related to the proving and verification time for all employed zk-SNARKs. As we show below for all three different applications and policies the proving time is in the order of seconds while the verification time in milliseconds. We report the average data across 10 runs.

ProveQual includes 8 checks as described in Figure 3, one of which is a membership proof for MT. This is the only variable for the generation of $\pi^{\tau}_{o_i}$ and we examine how the Merkle tree depth affects the proving and verification times. We present our findings for varying depths from 20 to 24 in Figure 7. Notably, to generate a proof for depth 23 a worker needs < 23 secs, which the requester verifies in ≈ 24 msecs.

AuthCalc embodies the implementation of AnsCalc. As such, we provide the implementation for the γ-Most Frequent (MF) and the Average (Avg) algorithms. The performance of AuthCalc depends

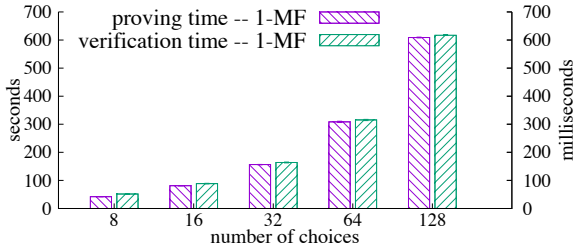

Figure 6: AuthCalc: Proving & Verification Time vs. number of choices for AnsCalc= 1-MF with 1024 workers.

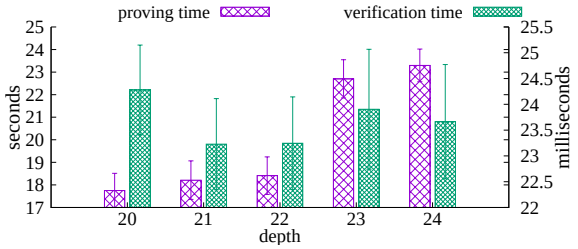

Figure 7: ProveQual: Proving & Verification Time vs. *MT* Depth.

AnsCalc	zk-SNARK	Proving (s)	Verification (ms)
1-MF	AuthQual	2.29	11.8
	AuthValue	1.93	11.2
3-MF	AuthQual	3.59	12.3
	AuthValue	3.25	11.7
Avg	AuthQual	2.91	12.1
	AuthValue	1.96	10.7

Table 4: AuthQual and AuthValue proving and verification time for different AnsCalc mechanisms.

on the number of *(i)* workers and *(ii)* choices. Figure 5 depicts the performance of both implementations (for $\gamma = \{1, 3\}$) with 4 choices and workers varying from 64 to 1024. Contrary, in Figure 6 we show the performance when fixing the number of workers to 1024 and varying the choices from 4 to 64. Crucially, the results confirm the practicality of our design, e.g., a requester using AVeCQ can output a proof pretending to be the most frequent answer, for 1K workers and 64 choices, in < 15 mins. We remark here that Avg is independent of the number of choices. Last, in all cases above, the verification time is < 0.7 secs.

AuthValue and AuthQual. The former produces a proof for the correctness of response w_i. For AnsCalc=1-MF we establish a^{τ}_i as correct if $a^{\tau}_i = a^{\tau}_{\phi}$, while for AnsCalc=Avg if it is within a P^{τ}-specific range around the final answer. Now, AuthQual produces a proof regarding the corresponding update of q_i based on the correctness of a^{τ}_i. Recall that the requester does not have access to the qualities in the plain. To update a worker's quality, the requester "adds" a Pedersen commitment of 0 or 1 to the worker-provided commitment appropriately. These zk-SNARKs have a constant number of constraints, and we provide the per-worker proof generation and verification times in Table 4.

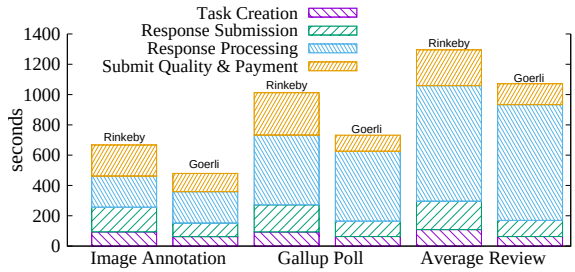

Figure 8: E2E completion time for different tasks with base fee 5 GWei and Priority fee 1 Gwei.

7.3 End-to-End (E2E) Run Time

To further demonstrate the practicality of AVeCQ, we measure its performance on popular crowdsourcing tasks [39, 62, 82] simulated on real-world datasets [26, 53, 75]. We measure each task's time from their deployment until completion.

Tasks. First, we consider Image Annotation [60], commonly used in crowdsourcing to generate datasets to train Machine Leaning (ML) models. Similar to [39], we consider a task such as identifying whether a given image contains a duck or not. We deploy $\texttt{CSTask}_{ML}$ for this task. Second, we consider the task of generating the Average Review of an online product through $\texttt{CSTask}_{Review}$ [39]. Last, we deploy $\texttt{CSTask}_{Gallup}$ to estimate public opinion through a Gallup poll. E.g. to determine the "COVID-19 fear level" for citizens living in major Spanish cities [53].

Datasets. For $\texttt{CSTask}_{ML}$, we use worker reports obtained using a real-world image annotation dataset, Duck [75]. Further, $\texttt{CSTask}_{ML}$ has AnsCalc=1-MF, $n^{\tau}_{th} = 39$, and $|A^{\tau}| = 2$. For $\texttt{CSTask}_{Gallup}$, the worker reports are from the real-world survey conducted by Pérez *et al.* [53]. Specifically, we have AnsCalc=3-MF, $|A^{\tau}| = 5$ and we subsample $n^{\tau}_{th} = 64$ worker reports from the $\approx$ 8K available. Last, $\texttt{CSTask}_{Review}$ acquires an average review, with individual worker reports taken from the Amazon review dataset [26]. More concretely, $\texttt{CSTask}_{Review}$ has AnsCalc=Avg, where each worker can report a rating from the set $\{1, 2, 3, 4, 5\}$, i.e., $|A^{\tau}| = 5$. We subsample $n^{\tau}_{th} = 128$ worker reports from the $\approx 1M$ available.

Results. Figure 8 depicts our results with $\delta = 1$ GWei. Depending on the task, a worker's response takes 93.34-108.52 secs (Rinkeby) and 62.36-63.12 secs (Goerli) to construct, submit, and get verified on-chain. Depending on the underlying crowdsourcing policy, the number of workers, and choices, to calculate the result and produce the corresponding proof the requester takes from 206 to 764 secs, while for uploading the quality updates and payments it takes 204-280 secs on Rinkeby and 106-140 secs on Goerli. Last, the quality verification phase takes $<$ 1 sec.

7.4 Comparing AVeCQ with Prior Works

We remark that AVeCQ outperforms state-of-the-art protocols [17, 41, 45, 46] in multiple metrics, despite incorporating worker quality and achieving stronger security properties. Below we compare wherever and however possible with such systems. We discuss initially E2E performance and then comparable individual aspects.

E2E Comparison. Surprisingly, very few prior works report the computational expenses or other overheads for an E2E execution of a specific task holistically. In fact, only [45] and [17] report such data. Table 5, includes a head-to-head comparison between our system and these works. As shown, AVeCQ outperforms [45] in time and [17] in gas required while achieving similar time efficiency.

Tasks	Image Annotation ($\lvert W\rvert = 39, \lvert A\rvert = 2$)	Average Review ($\lvert W\rvert = 128, \lvert A\rvert = 5$)
AVeCQ	Time: $<$ 8mins Gas: $<$ 19MWei	Time: $<$ 18mins Gas: $<$ 55MWei
E2E Comparable Blockchain-based Systems	ZebraLancer [45] Time: $>$ 7h* Gas: No data	Duan et al. [17] Time: 30min* Gas: $>$ 416MWei*

* denotes extrapolation of results.

Table 5: Completion time of AVeCQ (on Goerli) vs contemporary works in popular crowdsourcing tasks.

Gas Consumed. Task creation requires $\approx 4\times$ and response submission $\approx 7.4\times$ more gas in [46], compared to AVeCQ. Additionally, the deployment cost of the smart contracts in [17, 41] is comparable to the one of CSTask. Whereas, worker responses require ≈ 2 times more gas in [17]. Further, the authors of [41] perform the majority of their protocol operations on-chain, which would incur prohibitive gas (and monetary) costs for tasks with a high number of workers, based on a similar approach followed in [38].

Supporting Workers and Task Choices. Last, unlike [45, 46] we provide results for significantly greater number of workers and task choices. The authors in [45] provide results up to 11 workers and 2 choices, and only 4 workers in [46]. To the best of our knowledge, AVeCQ is the first anonymous crowdsourcing system measured against $>$ 1K workers and $>$ 100 choices (Figures 5 and 6).

8 DISCUSSION & CONCLUSION

Discussion. Using gold-standard tasks to evaluate the "trustworthiness" of a worker is quite popular [23, 25, 43, 81]. AVeCQ can support such tasks. In fact, for these types of tasks, all requester off-chain computations are almost non-existent since all workers know the final answer upon completion of the task and, therefore, can verify the correctness of their quality updates and payments. Therefore, AVeCQ operates in a richer setting functionality-wise, and specifically for this subset of cases, it is even more efficient. On a different note, there indeed exist quality/reputation systems based on more complex algebraic relations than additions (e.g., products [49] or Gompertz function [33]). We acknowledge this and identify as an interesting future direction the construction of an even more general protocol that can incorporate sophisticated quality scores inexpressible via homomorphic commitments (e.g., [33, 49, 64]). Last, AVeCQ is task/policy/blockchain agnostic at its core. The only requirement is that the policy can be expressed as a circuit and thus AnsCalc, QualCalc, and PaymCalc as zk-SNARKs.

Conclusion. In this work, we proposed AVeCQ, the first anonymous crowdsourcing system with verifiable worker qualities. AVeCQ leverages a fusion of cryptographic techniques and is built atop a blockchain that supports smart contracts. Moreover, we demonstrated via extensive experimentation that our system is deployable in real-world settings. Additionally, increases in the number of workers/choices for popular task policies do not impact the performance of AVeCQ. In conclusion, AVeCQ outperforms state-of-the-art and guarantees stronger security and privacy properties.

REFERENCES

[1] [n.d.]. Amazon Mechanical Turk. mturk.com.
[2] [n.d.]. Ethereum Gas Tracker. etherscan.io/gastracker.
[3] [n.d.]. Goerli Test Network. goerli.net/.
[4] [n.d.]. Microwork. microwork.app.
[5] [n.d.]. Qmarkers: Collective Intelligent Soltuions. qmarkets.net.
[6] [n.d.]. Rinkeby: Network Dashboard. rinekby.io.
[7] [n.d.]. Rinkeby Testnet Explorer. rinkeby.etherscan.io.
[8] Daniel J. Bernstein, Niels Duif, Tanja Lange, Peter Schwabe, and Bo-Yin Yang. 2011. High-Speed High-Security Signatures. In *CHES*. 124–142.
[9] Vincent Bindschaedler, Reza Shokri, and Carl A Gunter. 2017. Plausible deniability for privacy-preserving data synthesis. *VLDB Endowment* (2017), 481–492.
[10] Andrea Biondo, Mauro Conti, Lucas Davi, Tommaso Frassetto, and Ahmad-Reza Sadeghi. 2018. The Guard's Dilemma: Efficient Code-Reuse Attacks Against Intel SGX. In *USENIX*. 1213–1227.
[11] Ferdinand Brasser, Urs Müller, Alexandra Dmitrienko, Kari Kostiainen, Srdjan Capkun, and Ahmad-Reza Sadeghi. 2017. Software Grand Exposure: SGX Cache Attacks Are Practical. In *USENIX WOOT*.
[12] Sonja Buchegger and Jean-Yves Le Boudec. 2003. *A robust reputation system for mobile ad-hoc networks.* Technical Report.
[13] Vitalik Buterin, Eric Conner, Rick Dudley, Matthew Slipper, Ian Norden, and Abdelhamid Bakhta. 2019. EIP-1559: Fee market change for ETH 1.0 chain. eips.ethereum.org/EIPS/eip-1559.
[14] Dimitris Chatzopoulos, Mahdieh Ahmadi, Sokol Kosta, and Pan Hui. 2018. FlopCoin: A Cryptocurrency for Computation Offloading. *IEEE Transactions on Mobile Computing* 17, 5 (2018), 1062–1075. https://doi.org/10.1109/TMC.2017.2748133
[15] Ivan Bjerre Damgård. 1989. A design principle for hash functions. In *CRYPTO*. 416–427.
[16] Chris Dannen. 2017. *Introducing Ethereum and solidity*. Vol. 318.
[17] Huayi Duan, Yifeng Zheng, Yuefeng Du, Anxin Zhou, Cong Wang, and Man Ho Au. [n.d.]. Aggregating Crowd Wisdom via Blockchain: A Private, Correct, and Robust Realization. In *IEEE PerCom 2019*. 43–52.
[18] Jacob Eberhardt and Stefan Tai. 2018. Zokrates-scalable privacy-preserving off-chain computations. In *IEEE iThings and IEEE GreenCom and IEEE CPSCom and IEEE SmartData*. 1084–1091.
[19] Kevin Emery, Taylor Sallee, and Qi Han. 2015. Worker Selection for Reliably Crowdsourcing Location-Dependent Tasks. In *MobiCASE*.
[20] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *2014 ACM SIGSAC CCS*. 1054–1067.
[21] Boi Faltings, Radu Jurca, Pearl Pu, and Bao Duy Tran. 2014. Incentives to Counter Bias in Human Computation. In *HCOMP*. 59–66.
[22] Sheng Gao, Xiuhua Chen, Jianming Zhu, Xuewen Dong, and Jianfeng Ma. 2021. TrustWorker: A Trustworthy and Privacy-Preserving Worker Selection Scheme for Blockchain-based Crowdsensing. *IEEE TSC* (2021).
[23] Xi Alice Gao, James R. Wright, and Kevin Leyton-Brown. 2019. Incentivizing evaluation with peer prediction and limited access to ground truth. *Artificial Intelligence* 275 (2019), 618–638.
[24] Stylianos Gisdakis, Thanassis Giannetsos, and Panagiotis Papadimitratos. 2016. Security, Privacy, and Incentive Provision for Mobile Crowd Sensing Systems. *IEEE Internet of Things Journal* 3, 5 (2016), 839–853.
[25] Naman Goel and Boi Faltings. 2019. Deep bayesian trust: A dominant and fair incentive mechanism for crowd. In *AAAI*, Vol. 33. 1996–2003.
[26] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *WWW*. 507–517.
[27] Ethan Heilman, Leen Alshenibr, Foteini Baldimtsi, Alessandra Scafuro, and Sharon Goldberg. 2017. TumbleBit: An Untrusted Bitcoin-Compatible Anonymous Payment Hub. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017, San Diego, California, USA, February 26 - March 1, 2017*. The Internet Society. https://www.ndss-symposium.org/ndss2017/ndss-2017-programme/tumblebit-untrusted-bitcoin-compatible-anonymous-payment-hub/
[28] Daira Hopwood, Sean Bowe, Taylor Hornby, and Nathan Wilcox. 2016. Zcash protocol specification. *GitHub: San Francisco, CA, USA* (2016).
[29] Jeff Howe et al. 2006. The rise of crowdsourcing. *Wired magazine* 14, 6 (2006), 1–4.
[30] Shuang Hu, Lin Hou, Gongliang Chen, Jian Weng, and Jianhua Li. [n.d.]. Reputation-based distributed knowledge sharing system in blockchain. In *EAI MobiQuitous 2018*. 476–481.
[31] Mengdi Huai, Di Wang, Chenglin Miao, Jinhui Xu, and Aidong Zhang. 2019. Privacy-aware Synthesizing for Crowdsourced Data. In *IJCAI*. 2542–2548.
[32] Hyo Jin Jo and Wonsuk Choi. 2019. BPRF: Blockchain-based privacy-preserving reputation framework for participatory sensing systems. *Plos one* 14, 12 (2019), e0225688.
[33] Samhita Kanaparthy, Sankarshan Damle, and Sujit Gujar. 2022. REFORM: Reputation Based Fair and Temporal Reward Framework for Crowdsourcing. In *AAMAS*. 1648–1650.
[34] Thivya Kandappu, Arik Friedman, Vijay Sivaraman, and Roksana Boreli. 2015. Privacy in crowdsourced platforms. In *Privacy in a Digital, Networked World*. Springer, 57–84.
[35] David R Karger, Sewoong Oh, and Devavrat Shah. 2011. Iterative learning for reliable crowdsourcing systems. NeurIPS, 1953—-1961.
[36] Neal Koblitz. 1987. Elliptic curve cryptosystems. *Mathematics of computation* 48, 177 (1987), 203–209.
[37] Ahmed Kosba, Andrew Miller, Elaine Shi, Zikai Wen, and Charalampos Papamanthou. 2016. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In *2016 IEEE S&P*. IEEE, 839–858.
[38] Vlasis Koutsos, Dimitrios Papadopoulos, Dimitris Chatzopoulos, Sasu Tarkoma, and Pan Hui. 2021. Agora: A Privacy-Aware Data Marketplace. *IEEE TDSC* (2021), 1–1. https://doi.org/10.1109/TDSC.2021.3105099
[39] Evgeny Krivosheev, Siarhei Bykau, Fabio Casati, and Sunil Prabhakar. 2020. Detecting and preventing confused labels in crowdsourced data. *VLDB Endowment* 13, 12 (2020), 2522–2535.
[40] Nicolas Lambert and Yoav Shoham. 2009. Eliciting truthful answers to multiple-choice questions. In *ACM EC*. 109–118.
[41] Ming Li, Jian Weng, Anjia Yang, Wei Lu, Yue Zhang, Lin Hou, Jia-Nan Liu, Yang Xiang, and Robert H Deng. [n.d.]. CrowdBC: A blockchain-based decentralized framework for crowdsourcing. *TPDS 2018* 30, 6 ([n. d.]), 1251–1266.
[42] Qinghua Li and Guohong Cao. 2014. Providing Efficient Privacy-Aware Incentives for Mobile Sensing. In *2014 IEEE 34th International Conference on Distributed Computing Systems*. 208–217.
[43] Yihuai Liang, Yan Li, and Byeong-Seok Shin. 2022. Decentralized Crowdsourcing for Human Intelligence Tasks with Efficient On-Chain Cost. In *VLDB*. 1875–1888.
[44] D. Liu, A. Alahmadi, J. Ni, X. Lin, and X. Shen. 2019. Anonymous Reputation System for IIoT-Enabled Retail Marketing Atop PoS Blockchain. *IEEE TII* 15, 6 (2019), 3527–3537.
[45] Yuan Lu, Qiang Tang, and Guiling Wang. [n.d.]. Zebralancer: Private and anonymous crowdsourcing system atop open blockchain. In *IEEE ICDCS 2018*. 853–865.
[46] Yuan Lu, Qiang Tang, and Guiling Wang. 2020. Dragoon: Private decentralized hits made practical. In *IEEE ICDCS*. 910–920.
[47] Yuan Luo and Nicholas R. Jennings. 2018. A Differential Privacy Mechanism with Network Effects for Crowdsourcing Systems. In *AAMAS*. 1998–2000.
[48] Ralph C Merkle. 1987. A digital signature based on a conventional encryption function. In *Conference on the theory and application of cryptographic techniques*. 369–378.
[49] Moin Hussain Moti, Dimitris Chatzopoulos, Pan Hui, and Sujit Gujar. 2019. FaRM: Fair Reward Mechanism for Information Aggregation in Spontaneous Localized Settings. In *IJCAI*. 506–512.
[50] David Oleson, Alexander Sorokin, Greg Laughlin, Vaughn Hester, John Le, and Lukas Biewald. 2011. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. In *HCOMP*. 43–48.
[51] Torben Pryds Pedersen. 1991. Non-interactive and information-theoretic secure verifiable secret sharing. In *CRYPTO*. 129–140.
[52] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* (2014), 1023–1031.
[53] Virgilio Pérez, Cristina Aybar, and Jose M Pavía. 2022. Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain. *Data in Brief* 40 (2022), 107700.
[54] Drazen Prelec. 2004. A Bayesian Truth Serum for Subjective Data. *Science (New York, N.Y.)* 306 (11 2004), 462–466. https://doi.org/10.1126/science.1102081
[55] Goran Radanovic, Boi Faltings, and Radu Jurca. 2016. Incentives for effort in crowdsourcing using the peer truth serum. *ACM TIST* (2016), 1–28.
[56] Christian Reitwiessner. 2017. EIP 196: Precompiled contracts for addition and scalar multiplication on the elliptic curve alt_bn128. *Ethereum Improvement Proposals* 196 (2017).
[57] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, et al. [n.d.]. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *CHI 2015*. 1621–1630.
[58] Eli Ben Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. 2014. Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE S&P*. 459–474.
[59] Leonard J Savage. 1971. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* 66, 336 (1971), 783–801.
[60] Nihar Bhadresh Shah and Dengyong Zhou. 2015. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. *NeurIPS* 28 (2015), 1–9.
[61] Victor S. Sheng, Foster Provost, and Panagiotis G. Ipeirotis. 2008. Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers *(ACM KDD)*. 614 – 622.
[62] Haipei Sun, Boxiang Dong, Hui Wang, Ting Yu, and Zhan Qin. 2018. Truth inference on sparse crowdsourcing data with local differential privacy. In *IEEE Big Data*. IEEE, 488–497.
[63] Cristian Tanas, Sergi Delgado-Segura, and Jordi Herrera-Joancomartí. [n.d.]. An Integrated Reward and Reputation Mechanism for MCS Preserving Users'

Privacy. In *DPM/QASA@ESORICS 2015*. 83–99.
[64] Y. Tang, S. Tasnim, N. Pissinou, S. S. Iyengar, and A. Shahid. 2018. Reputation-Aware Data Fusion and Malicious Participant Detection in Mobile Crowdsensing. In *IEEE Big Data*. 4820–4828.
[65] Qian Tao, Yongxin Tong, Zimu Zhou, Yexuan Shi, Lei Chen, and Ke Xu. 2020. Differentially private online task assignment in spatial crowdsourcing: A tree-based approach. In *IEEE ICDE*. IEEE, 517–528.
[66] Hien To, Gabriel Ghinita, and Cyrus Shahabi. 2014. A framework for protecting worker location privacy in spatial crowdsourcing. *VLDB Endowment* (2014), 919–930.
[67] Hien To, Gabriel Ghinita, and Cyrus Shahabi. 2015. PrivGeoCrowd: A toolbox for studying private spatial crowdsourcing. In *IEEE ICDE*. IEEE, 1404–1407.
[68] Hongwei Wang, Song Guo, Jiannong Cao, and Minyi Guo. 2017. MeLoDy: A long-term dynamic quality-aware incentive mechanism for crowdsourcing. *IEEE Trans. on Parallel and Distributed Systems* 29, 4 (2017), 901–914.
[69] J. Wang, J. Tang, D. Yang, E. Wang, and G. Xue. 2016. Quality-Aware and Fine-Grained Incentive Mechanisms for Mobile Crowdsensing. In *IEEE ICDCS*. 354–363.
[70] Leye Wang, Gehua Qin, Dingqi Yang, Xiao Han, and Xiaojuan Ma. 2018. Geographic differential privacy for mobile crowd coverage maximization. In *AAAI*. 200–207.
[71] Leye Wang, Dingqi Yang, Xiao Han, Tianben Wang, Daqing Zhang, and Xiaojuan Ma. 2017. Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation. In *WWW*. 627–636.
[72] Leye Wang, Daqing Zhang, Dingqi Yang, Brian Y Lim, and Xiaojuan Ma. 2016. Differential location privacy for sparse mobile crowdsensing. In *IEEE ICDM*. 1257–1262.
[73] Yingjie Wang, Zhipeng Cai, Guisheng Yin, Yang Gao, Xiangrong Tong, and Guanying Wu. 2016. An incentive mechanism with privacy protection in mobile crowdsourcing systems. *Computer Networks* 102 (2016), 157–171.
[74] Yue Wang, Ke Wang, and Chunyan Miao. 2020. Truth Discovery against Strategic Sybil Attack in Crowdsourcing. In *KDD*. 95–104.
[75] Peter Welinder, Steve Branson, Pietro Perona, and Serge Belongie. 2010. The multidimensional wisdom of crowds. *NeurIPS* 23 (2010), 2424—-2432.
[76] Jens Witkowski and David Parkes. 2012. A Robust Bayesian Truth Serum for Small Populations. In *AAAI*. 1492–1498.
[77] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum yellow paper* 151 (2014), 1–32.
[78] Xingfu Yan, Wing WY Ng, Biao Zeng, Changlu Lin, Yuxian Liu, Lu Lu, and Ying Gao. 2021. Verifiable, reliable, and privacy-preserving data aggregation in fog-assisted mobile crowdsensing. *IEEE Internet of Things Journal* (2021).
[79] Xiang Zhang, Guoliang Xue, Ruozhou Yu, Dejun Yang, and Jian Tang. 2014. You better be honest: Discouraging free-riding and false-reporting in mobile crowdsourcing. In *2014 IEEE GLOBECOM*. 4971–4976.
[80] Xinglin Zhang, Zheng Yang, Wei Sun, Yunhao Liu, Shaohua Tang, Kai Xing, and Xufei Mao. 2016. Incentives for Mobile Crowd Sensing: A Survey. *IEEE Communications Surveys Tutorials* 18, 1 (2016), 54–67. https://doi.org/10.1109/COMST.2015.2415528
[81] Bowen Zhao, Shaohua Tang, Ximeng Liu, and Xinglin Zhang. 2021. PACE: Privacy-Preserving and Quality-Aware Incentive Mechanism for Mobile Crowdsensing. *IEEE TMC* 20, 5, 1924–1939.
[82] Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. 2017. Truth inference in crowdsourcing: Is the problem solved? *VLDB Endowment* 10, 5 (2017), 541–552.
[83] Saide Zhu, Zhipeng Cai, Huafu Hu, Yingshu Li, and Wei Li. 2019. zkCrowd: a hybrid blockchain-based crowdsourcing platform. *IEEE Transactions on Industrial Informatics* 16, 6 (2019), 4196–4205.

A SECURITY PROPERTIES OF AVeCQ

A.1 Anonymity (with Unlinkability)

The goal is to define the property so that no entity can establish a connection between a worker's identity and its responses across tasks. We therefore need to take a closer look to the response o_i. Remember that a worker w_i submits o_i^τ to a task τ in the following form: $o_i^\tau = \langle E(pk_R, a_i^\tau; \cdot), E(pk_R, pa_i^\tau; \cdot), \mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1} + r_{\star,i}^\tau), E(pk_R, r_{k,i}^\tau; \cdot), H\big(\mathsf{Com}(q_i^{\tau-1}, r_{c,i}^{\tau-1}), m_i\big), \pi_{o_i}^\tau \rangle$.

First, we look at the public address pa_i: if is not unique across tasks, then any requester (or participant in general) can connect the same pa_i with two discrete responses and thus identify-link a worker across two tasks. Notably, this is the case regardless of whether the uploading mechanism is anonymous or not. For example, in Ethereum a worker may select to respond to different tasks from equally different addresses. However, this measure alone is not sufficient. In our system's case, by including pa_i encrypted in the response tuple, we ensure that only the requester can identify any connection between the uploading mechanism and the public addresses that receive compensation after the task has been concluded. This, in turn, translates into the fact that *any* other entity that is even monitoring *all* blockchain transactions may connect each worker across *at most* two tasks. To avoid even such a case though, multiple approaches have already been proposed in the literature, from anonymous tokens e.g., ZCash [28] to mixnets or tumblers [27]; techniques easy to plug in AVeCQ.

We now consider which other element of the response can be used non-trivially to link a worker across two tasks. We observe that the quality scores can "reveal" the identity of a worker across two tasks. If a worker submits its quality in the response as a plaintext then any participant can possibly track them across tasks e.g., workers with high or low qualities. However, even if they are provided in a hiding manner i.e., as commitments in AVeCQ, recall that it is actually the requester that updates the quality scores of the participating workers and then commits to them. Therefore, if qualities are used exactly as provided by the last requester then back-to-back responses can be linked across these two tasks, forming in fact the full chain or responses across *all* tasks. To avoid this we require from workers to re-randomize the quality commitments before participating in the next task.

To prove that AVeCQ satisfies Anonymity we rely on the hiding property of Com, the collision-resistance of H, and the zero-knowledge property of ProveQual, assuming that no non-trivial information about q_i is leaked to the requester or anyone else during response submission or processing. We refer to trivial leakage as information that one can decipher from the publicly available task policy (e.g., q_i surpasses a certain q_{th}^τ). We design a game $\mathcal{G}_{Anon}$ to formally capture anonymity as a property. At a high level, we state that a crowdsourcing system is anonymous if for any two distinct tasks τ_i, τ_j with corresponding worker sets $\mathcal{W}_i, \mathcal{W}_j$, no entity can non-trivially distinguish whether $\exists\, w_\star$ s.t. $w_\star \in \mathcal{W}_i$ and $w_\star \in \mathcal{W}_j$.

A.2 Free-rider Resistance

For this property we try and capture the following adversarial behavior: A worker whose goal is to utilize another worker's answer and/or quality and get compensated. We present and analyze all three cases below.

First, workers might try to elude task execution or inflate total worker rewards[9] by submitting other workers' responses as their own. Generally in such a case the requester would face the challenge of deciding which of the duplicate tuples was submitted first/legitimately. However, since AVeCQ utilizes a blockchain all submitted responses are timestamped, meaning that even if a worker copies all parts of a response and only changes the public address part (or even submit the same response entirely), the requester can disregard any "duplicate" records. Thus, such behavior is counterincentivized as workers who pawn someone else's response as their own *(i)* have to expend resources for uploading it on-chain and *(ii) deterministically* will get no reward.

Second, workers might try to use someone else's quality to pass a specified threshold. In AVeCQ, when responding to a task, w_i ties specifically its q_i to the underlying m_i of $cert_i$. Then, each worker produces a proof of validity for all hashes, commitments, and encryptions in the submit-response tuple using ProveQual. From the soundness property of the employed zk-SNARK and the hiding property of the commitment scheme, no polynomially bound worker can produce a convincing proof without a witnesses and from the zero-knowledge property of ProveQual no polynomially-bound worker can extract any underlying witness from $\pi_{o_i}^\tau$.

Third, to ensure no worker can utilize the encrypted-answer part of another worker's response and use it with its own quality commitments we rely on the soundness and zero-knowledge property of ProveQual. Similarly to the above, no polynomially-bound worker can extract another worker's answer a_i from its response o_i. We design a game $\mathcal{G}_{FRR}$ to formally capture this property. At a high level, we state that a crowdsourcing system is secure against free-riding attacks if the adversary can submit a response o_i containing *any* element from another worker's response o_i' and can distinguish between getting the legitimate compensation based on the answer included in o_i or the minimum possible compensation.

A.3 Policy-Verifiable Correctness

No requester can produce convincing fabricated $a_\phi^{\tau,\prime}$ and qualities. Additionally, no worker-provided response that contains an answer $a_i \notin A^\tau$ is included in the a_ϕ^τ calculation. To prove this, we rely on the soundness of the zk-SNARKs' outputs: $\pi_\phi^\tau \leftarrow$ AuthCalc$(\cdot)$, $\pi_{q_i} \leftarrow$ AuthQual$(\cdot)$, and $\pi_{a_i}^\tau \leftarrow$ AuthValue$(\cdot)$. The trusted RA setups all SNARKS for policy P^τ. Thus, no requester can produce a valid proof that pertains to a fabricated result $a_\phi^{\tau,\prime} \neq a_\phi^\tau$ or qualities for a different policy $P^{\tau,\prime}$ without breaking the underlying zk-SNARK soundness property.

A.4 Other Properties

Sybil-Attack Resistance. These attacks correspond to entities forging identities to participate in tasks. For instance, a worker with a "low" quality may prefer to generate fresh identities to have higher chances of clearing quality thresholds and being able to participate in future tasks. Moreover, workers might attempt to participate in

[9] E.g., a worker who cannot clear the quality threshold for a task can collude with another worker, submit a duplicate response, and split the reward.

a task multiple times, under different identities, and reap rewards almost arbitrarily. On top of that, especially in tasks where the final answer is not known a-priori, a worker that participates arbitrarily-many times in a task can launch even more sophisticated attacks (see Section A.5). The severity of Sybil attacks in crowdsourcing systems has been studied extensively and we rely on a trusted RA to combat them, similarly to prior works [43, 45, 46]. Recall that AVeCQ includes a registration phase where every worker provides a secret message m_i and acquires a participation certificate in the form of an EdDSA signature $cert_i$ through the RA. To submit o_i, each worker has to generate a ProvQual proof, which includes a verification of $cert_i$ based on m_i. From modeling the RA as trusted and the soundness property of ProveQual, no polynomially-bound worker can produce a convincing proof without the respective witnesses.

Payment and Quality Deprivation Resistance. No requester can avoid paying. CSTask method PayWorker handles the payments of workers during the quality verification phase. Therefore, rightful reimbursements are allocated to workers, assuming an honest majority in the on-chain validators. Any worker w_i who submits a correct answer a_i^τ and yet is not paid for participating in τ, during the protest period, can contact the RA with the evidence $\pi_{a_i}^\tau, o_i^\tau$. Upon verifying the proof, the RA can confiscate the requester funds, pay the worker, and impose added penalties (similarly, for when a requester does not upload on-chain a quality update).

A.5 Miscellaneous Attacks

Below we analyze attacks that our crowdsourcing and threat model (Section 4) allow and possible mitigation techniques.

Final-Answer Skewing. First, since we allow arbitrary collusions between the entities and our system imposes no restrictions to the task policy, the following attack is enabled. Consider our running-example task where $|\mathcal{W}^{\text{Yes}}| = |\mathcal{W}^{\text{No}}| + 1$. An adversarial requester can skew the final answer if he colludes with even just two additional workers. Interestingly, this type of attack in combination with certain task policies can even result in the requester giving out less rewards totally!

The following example is illuminating: Consider the case where the final answer is calculated as the average of the workers numerical responses and the qualities/rewards are calculated based on a proximity deviation between the worker's answer and the final answer. In this case, the requester can collude with even just one worker who just needs to purposefully submit an "overshot" answer, affecting the final answer enough to reduce total expended rewards. Our system does not safeguard workers against such game-theoretic attacks, which is in line with the broad mechanism design [21, 33, 49, 54, 55, 76] and anonymous crowdsourcing [17, 45] literature. Additionally, in tasks where the final answer of a task is calculated as a function of participating worker responses, workers may opt to misreport their answer to try and "guess" the final answer and get rewarded. Popular approaches to avoid such attacks is to impose constant rewards for all participating workers or enable tasks with only publicly available ground truth [40, 59]. In AVeCQ we adopt a more expressing model concerning the task policy, which includes the above approaches, but is not restricted to these.

Quality Inflation. Another possible (and rather subtle) threat stems from the fact that in AVeCQ workers verify just their own quality updates. In fact, if a worker realizes that its quality is *lower* than what it should be or that the requester-provided proofs do not pass verification, it is incentivized to protest and correct the wrongdoing. However, if the updated quality is *higher*, then the worker is incentivized to avoid protesting! This in turn, enables requesters to collude with workers to raise arbitrarily their qualities. Previous blockchain-based works adopt on-chain verification to avoid similar issues [45, 46]. We adopt a different approach; all proofs and responses are uploaded on-chain, but all verifications happen off-chain. Thus, to combat such behavior we can pair AVeCQ with a "bounty-hunter" protocol, where blockchain participants are incentivized to verify all requester-provided proofs and reap additional rewards upon discovering rejecting ones.

We denote by $Adv_x^{\mathcal{G}}(\mathcal{A})$ the advantage that the adversary $\mathcal{A}$ has in winning the game $\mathcal{G}_x$. Similarly we denote by $Adv^{\text{C-Hiding}}(\mathcal{A})$ the advantage $\mathcal{A}$ has in breaking the hiding property of the commitment scheme C, by $Adv^{H\text{-CR}}(\mathcal{A})$ breaking the collision-resistance property of H, and by $Adv^{\text{ProveQual-ZK}}(\mathcal{A})$ breaking the zero-knowledge property of ProveQual.

B ZK-SNARKS SPECIFICATIONS FOR AVecQ

Figure 3 includes all the parameters and formal languages supported by the SNARKs used in AVecQ.

ProveQual. For a task τ, each worker w_i generates a proof $\pi_{o_i}^\tau$ attesting to the correctness of its quality q_i^τ using ProveQual. The statement $\vec{x}_i$ comprises the MT root, requester, and RA's public keys, the re-randomized commitment of w_i's current quality, the quality "tag" and encryption of w_i's answer and address for reimbursement. The corresponding witness $\vec{\omega}_i$ consists of w_i's certificate, its secret identifier, quality and corresponding randomness in plaintext, quality commitment, plaintext answer, and public address followed by the MT path. ProveQual's language-specific checks prove the correctness of $\vec{x}_i$ while simultaneously guarding w_i's response against free-rider attacks. As mentioned in Section 5.2, these checks include certificate verification (EdDSAver), quality verification (QualVer), correctness of the answer and address encryptions (ValidEnc), quality "tag" verification (HashComVer), MT membership proof (MTPathVer), and any task-specific check with TaskVer.

We can trivially see that these checks together ensure the validity of $\vec{x}_i$ in ProveQual. We also remark that ProveQual assists in safeguarding AVeCQ against free-rider attacks. By including the correctness check for w_i's reimbursement address in ProveQual, we ensure that an adversary copying w_i's response cannot change the reimbursement address (without breaking the SNARK's soundness property). As such, no adversary has an incentive to launch free-riding attacks. Lastly, with Figure 7, we also highlight the efficiency of ProveQual's design. The proving time for the requester, with > 1M responses (or depth > 20), is 18-25 seconds.

Workers use ProveQual to generate proofs for their responses. In comparison, a task's requester uses AuthCalc, AuthQual, and AuthValue to attest to the correctness of the final answer, a worker's quality and its proximity to the final answer, respectively.

AuthCalc. Each task's requester uses AuthCalc to generate the proof $\pi^{\tau}_{a_\phi}$ to attest the correctness of the final answer a^{τ}_{ϕ}. The statement $\vec{x}_{AC}$ includes the encryption of a^{τ}_{ϕ}, the set of all encrypted responses and the public key pk_R; while the witness $\vec{\omega}_{AC}$ comprises the requester's secret key sk_R. Each task's policy is hard-coded in the SNARK. The language makes up the following two checks: (i) ValidKeyPair, which ensures that (pk_R, sk_R) are valid key pair and (ii) FinAnsVer, which checks if the final answer is correctly computed given the set of all encrypted responses.

Figure 5 and Figure 6 shows the practicality of AuthCalc. E.g., the SNARK can generate proofs for popular crowdsourcing policies for 1K workers and 64 choices in < 15 minutes.

AuthQual and AuthValue. The requester uses these SNARKs to generate the proofs $\pi^{\tau}_{q_i}$ and $\pi^{\tau}_{a_i}$. These attest to the correctness of w_i's updated quality and the proximity of their answer with the final answer. Note that the quality update rule and worker answer policies are hard-coded in the SNARK. AuthValue's statement $\vec{x}_{AV}$ includes encryption of the final and w_i's answer and requester public key pk_R, while AuthQual's statement $\vec{x}_{AQ}$ additionally includes the commitments of the old and updated qualities. Both the witnesses, $\vec{\omega}_{AV}$ and $\vec{\omega}_{AQ}$, comprise the secret key sk_R. The language for AuthValue comprises checks for the key-pair validity (ValidKeyPair) and the correctness of w_i's answer with respect to a^{τ}_{ϕ} (EqCheck). Likewise, AuthQual's language includes ValidKeyPair and NewQual which checks the quality update given w_i's answer, a^{τ}_{ϕ} and the old and updated quality commitments as inputs.

Notably, from Table 4, both these SNARKs are efficient in proving and verification times for popular crowdsourcing policies. Depending on the policy, the proving time ranges from 2.9-3.6 seconds, while the verification takes < 12.5 milliseconds.