Fait Accompli Committee Selection

Improving the size-security tradeoff of stake-based committees

Fait Accompli Committee Selection: Improving the Size-Security Tradeoff of Stake-Based Committees

Peter Gaži
IOG
peter.gazi@iohk.io

Aggelos Kiayias
University of Edinburgh & IOG
aggelos.kiayias@ed.ac.uk

Alexander Russell
University of Connecticut & IOG
acr@uconn.edu

ABSTRACT

We study the problem of *committee selection* in the context of proof-of-stake consensus mechanisms or distributed ledgers. These settings determine a family of participating parties—each of which has been assigned a non-negative "stake"—and are subject to an adversary that may corrupt a subset of the parties. The challenge is to select a committee of participants that accurately reflects the proportion of corrupt and honest parties, as measured by stake, in the full population. The trade-off between committee size and the probability of selecting a committee that over-represents the corrupt parties is a fundamental factor in both security and efficiency of proof-of-stake consensus, as well as committee-run layer-two protocols.

We propose and analyze several new committee selection schemes that improve upon existing techniques by adopting low-variance assignment of certain committee members that hold significant stake. These schemes provide notable improvements to the size–security trade-off arising from the stake distributions of many deployed ledgers.

CCS CONCEPTS

• **Mathematics of computing** → **Probabilistic algorithms**; • **Computing methodologies** → **Distributed algorithms**.

KEYWORDS

Consensus; Delegation; Distributed ledgers; Committee selection

ACM Reference Format:
Peter Gaži, Aggelos Kiayias, and Alexander Russell. 2023. *Fait Accompli* Committee Selection: Improving the Size-Security Tradeoff of Stake-Based Committees. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23), November 26–30, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3576915.3623194

1 INTRODUCTION

We study the problem of *committee selection* in the context of proof-of-stake consensus mechanisms or distributed ledgers. These settings determine a family P of participating parties, with each $p \in P$ assigned a non-negative *stake* $\mathcal{S}(p)$. An unknown subset $A \subset P$ of the parties have been corrupted by an adversary, with the only constraint that a strict minority of total stake is corrupted, which is to say that

$$\sum_{p \in A} \mathcal{S}(p) \leq \alpha \cdot \sum_{p \in P} \mathcal{S}(p)\,, \tag{1}$$

for a parameter $\alpha < 1/2$. Committee selection is the problem of choosing a committee $C \subset P$ with the property that less than a $(\alpha+\epsilon)$-fraction of its members are corrupt, for a (related) parameter $\epsilon \in [0, 1-\alpha)$. Intuitively, such a committee permits a small number of parties to inherit—with small distortion ϵ—the desirable adversarial proportionality of the entire population, which can provide important efficiency or algorithmic advantages.

Committee selection is a natural—and often fundamental—aspect of proof-of-stake consensus mechanisms, with various settings of interest placing differing demands on α, ϵ, the size of the committee, and the probability of failure. In particular, "Iterated BFT" mechanisms operate by directly electing sequences of such committees; Algorand [5], for example, falls into this category, requires $\alpha + \epsilon < 1/3$, and demands committee sizes that scale appropriately with $1/\epsilon$. Committee selection likewise occurs, in some cases indirectly, in "Nakamoto-style" proof-of-stake consensus protocols; the Ouroboros blockchains [6, 11], for example, implicitly elect such committees in each epoch, require $\alpha + \epsilon < 1/2$, and demand epoch lengths that scale appropriately with $1/\epsilon$. Committee selection is also instrumental in common "layer two" infrastructure that is bootstrapped by a distributed ledger. In these cases, a committee is selected according to a stake distribution determined by an underlying distributed ledger and is then pressed into service for external purposes. Examples include maintaining a secondary blockchain (a "sidechain"), bridge, or payment or state channel, tallying and certifying an off-chain governance vote, and providing oracle services to on-chain smart contracts—we refer to these tasks collectively as *committee-run layer-two protocols*. In all these settings, one must balance the efficiency advantages of small committees against the security penalties: in particular, smaller committees increase the probability of disproportionate representation of adversarial parties. In the examples mentioned above the relationship between committee size and security (that is, the probability of such disproportionate representation) is of first order importance for security and efficiency.

There are two standard approaches to committee selection: the first calls for drawing n members—independently and proportionally to stake—for a suitably selected n; the second calls for independently including each participant (perhaps with multiplicity) with a probability that scales with its stake so as to yield a committee of expected size n. In this article, we show that when there are parties holding approximately a $1/n$ fraction of stake, committee selection can exploit this to provide an improved relationship between committee size and failure probability. Applying our techniques

CCS '23, November 26–30, 2023, Copenhagen, Denmark

ACM ISBN 979-8-4007-0050-7/23/11.
https://doi.org/10.1145/3576915.3623194

to the concrete stake distributions arising in currently deployed proof-of-stake ledgers, we observe striking improvements in the committee size required in order to achieve various error thresholds of interest. Considering Ethereum, for example, our approach leads to committees of size 300 with better security characteristics than existing approaches achieve even with size 1000. We compare some of the schemes we develop against classical techniques in Figure 1.

Conventional committee selection. We assume throughout that the adversary determines the set A of corrupt parties with full knowledge of the procedure that will be used to select the committee. If the procedure is deterministic, the adversary may select A with foresight so as to corrupt certain committee members with certainty; if, furthermore, the aggregate stake assigned to the committee members is a small fraction of total stake, the entire committee can be corrupted outright. Incorporating randomness in the procedure is the ready countermeasure to such "precognition attacks," and in this context the distribution obtained by selecting parties with probability proportional to stake plays a central role. For this reason, we assume throughout that stake assignments are scaled so that $\sum_{p\in P} \mathcal{S}(p) = 1$ and refer to $\mathcal{S}$ as the "stake distribution." Observe that when a committee member $c \in P$ is selected according to the distribution $\mathcal{S}$,

$$\Pr[c \text{ is corrupt}] = \Pr[c \in A] \le \alpha$$

for any adversarial choice of A so long as it satisfies the corruption budget of (1). This motivates one of the standard approaches to committee selection, which defines the committee C to consist of n parties $c_1, \ldots, c_n$ independently chosen according to $\mathcal{S}$ for a suitably large n; we denote this method IID. (Note that this process defines a multiset, in general, as individual parties may be drawn multiple times by this process.) It follows that the expected fraction of the resulting committee that is corrupt is no more than α, where committee members are weighted according to the number of times they have been selected. Standard Chernoff–Hoeffding bounds for the tails of the binomial distribution can then be applied to estimate the probability of a significant deviation: For example, a committee of size n will contain $(\alpha + \epsilon)n$ corrupt parties—overrepresenting the corrupt fraction of players by ϵn—with probability

$$\delta = \exp(-\Omega(n\epsilon^2)) \,. \tag{2}$$

Returning to the practical cases highlighted above, δ is a critical quantity that features prominently in the final security guarantees of the protocol while n features prominently in the efficiency of the protocol.

Motivating fait accompli committee selection. The approach in this article is motivated by the fact that—in certain unusual cases—committee selection can, in fact, be carried out deterministically and with zero probability of failure. Consider, for example, the setting with n parties and the uniform stake distribution $\mathcal{S}(p) = 1/n$ for each $p \in P$. Then the committee consisting of P itself will clearly have no more than a α-fraction of corrupt players for any subset A satisfying (1). This example suggests that optimal committee selection should exploit structural properties of the stake distribution to reduce variance. Our main result realizes this intuition, showing how to account for the structure of $\mathcal{S}$ during committee selection in order to achieve an improved relationship between committee size and probability of failure. In general, these methods provide improved performance when $\max_{p\in P} \mathcal{S}(p) \gtrapprox 1/n$ and, specifically, provide improvements determined by the ability of functions that take values in the set $\frac{1}{n}\mathbb{N}_0 = \{0, 1/n, 2/n, \ldots\}$ to approximate the distribution $\mathcal{S}$ for suitable choices of n. In more detail, if there is a function $f \le \mathcal{S}$ taking values in the set $\frac{1}{n}\mathbb{N}_0$ that covers all but a τ portion of the probability mass of $\mathcal{S}$ then the resulting committee of size n will suffer from an ϵn overrepresentation of corrupt parties with probability

$$\delta = \exp(-\Omega(n\epsilon^2/\tau)) \tag{3}$$

(cf. (2)). It is interesting to observe that the ability of functions taking values in $\frac{1}{n}\mathbb{N}_0$ to approximate $\mathcal{S}$ obviously improves with n, which is to say that $\tau \to 0$ as n increases. In particular, the fundamental scaling above improves as $n \to \infty$. In this context, we will be interested in the smallest choice of n for which the resulting error δ is driven below a desired probability of failure.

A summary of the results. We begin in Section 3 by focusing on the *unweighted* setting outlined above, where the proportion of corrupt players in the final committee is measured by cardinality. We there describe and analyze two schemes (FA1 and FA2) that provide security that scales with the ability of various families of functions to approximate $\mathcal{S}$.

We also focus on a second family of "local sortition schemes." To appreciate their importance, recall the second classical approach mentioned in passing above: Participants are individually elected to the committee independently (and perhaps with multiplicity) depending on their stake. While this approach has some additional complexity—e.g., the size of the committee itself is a random variable—by introducing public-key cryptographic tools (verifiable random functions [13]) a public seed can be used by the participants to *privately and locally* determine whether they are a member of the committee. This design grants such schemes an important feature: security against adaptive adversaries that may choose which participants to corrupt based on the public seed. We introduce a local-sortition variant of FA1 that achieves such adaptive security and hence can serve as a drop-in replacement for standard local sortition in protocols where this property is desired.

In Section 4 we turn our attention to *weighted* schemes, in which each committee member is assigned a non-negative weight and the corrupt proportion is measured according to weight. We find that this additional degree of freedom has a rather remarkable effect on security: in particular, our weighted scheme wFA squarely outperforms its unweighted counterparts: in many settings of practical interest, it provides orders of magnitude improvement in the resulting security guarantee.

We remark that the applicability of weighted schemes in practice depends on the setting: in circumstances where the job of the committee is to support a straightforward casting of votes, weighted schemes appear sufficient. On the other hand, for settings that require sophisticated cryptographic aggregation of votes or shares, such as secret sharing or threshold signatures, weighting may be unattractive.

Intuition for the main results and analysis. The core idea of fait accompli committee selection is to identify portions of the stake distribution with total probability $1/n$ that have low minimum

entropy—for example, are all assigned to a single party—and use these to guide committee selection. To illustrate the basic idea, consider the following simple procedure $\mathsf{FA1}(P, \mathcal{S}, n)$, which selects a committee of size n for a stake distribution $\mathcal{S}$ on P.

$\mathsf{FA1}(P, \mathcal{S}, n)$:

If $\max_{p \in P} \mathcal{S}(p) \geq 1/n$**:** Let p^* be a party for which $\mathcal{S}(p^*) \geq 1/n$. Add p^* to the committee and define

$$\mathcal{S}'(p) = \frac{n}{n-1} \cdot \begin{cases} \mathcal{S}(p) - 1/n & \text{if } p = p^*, \\ \mathcal{S}(p) & \text{otherwise.} \end{cases}$$

It is easy to verify that $\mathcal{S}'$ is a probability distribution. The rest of the committee is determined recursively as

$$\mathsf{FA1}(P, \mathcal{S}', n-1).$$

If $\max_{p \in P} \mathcal{S}(p) < 1/n$**:** The committee is determined by independently drawing n members according to $\mathcal{S}$.

To return to the intuition above, note that when there is a party for which $\mathcal{S}(p) \geq 1/n$, the party is added to the committee and a $1/n$ slice of the distribution associated with this party is removed from consideration; indeed, the residual distribution $\mathcal{S}'$ arises precisely by conditioning on the complement of this slice. Thus this committee member "perfectly represents" this portion of the distribution regardless of the adversary's selection of A and such a selection generates zero variance in the final family of random variables that reflect the corrupt portion of the committee. In the other case, members are selected according to the distribution as usual, generating variance around the current mean. In fact, this simple procedure yields the improved bound (3).

To concretely illustrate the benefits of the above procedure FA1 compared to the conventional IID method, consider a toy example where the stake is distributed in the following way: a single party p_{big} controls $1/3$ of all stake, while the rest is uniformly distributed among a large number of tiny stakeholders. A committee of size $n = 30$ is to be chosen, and the adversary is capable of corrupting parties controlling $1/3$ of all stake in total (i.e., he may either corrupt p_{big}, or one half of all the tiny stakeholders). The goal of the adversary is to control at least $1/2$ of the selected committee; i.e., this corresponds to the case $\alpha = 1/3$ and $\epsilon = 1/6$. If IID is used to select the committee, it is easy to see that the adversary achieves his goal with probability

$$\Pr\left[B(30, 1/3) \geq 15\right] \approx 4\%\,, \tag{4}$$

where $B(n, p)$ is a random variable having binomial distribution with n independent trials, each successful with probability p. On the other hand, if FA1 is used for the committee selection, the party p_{big} is deterministically assigned 10 of the 30 committee seats, and the rest is distributed in the standard way to the minor stakeholders. In this situation, it is clearly disadvantageous for the adversary to corrupt p_{big}, as this would *certainly* prevent him from achieving his goal: his representation within the committee would be guaranteed to be a $1/3$ fraction. Corrupting the maximum allowed number of the minor stakeholders instead (amounting to $1/3$ of total stake, which is $1/2$ of all minor stakeholders), the probability that the adversary now achieves his goal is

$$\Pr\left[B(20, 1/2) \geq 15\right] \approx 2\%\,,$$

as this requires that at least 15 out of the remaining 20 seats assigned by the standard method would end up controlled by a corrupted party. Put differently, it is easy to verify that if FA1 is used in this scenario, a committee of size 21 is sufficient to keep the probability of adversarial success below (4). This already illustrates a concrete advantage of FA1 over IID.

While the above example is encouraging and, in general, it seems plausible that the "perfect representation" of a $1/n$ portion of stake by a single, deterministically assigned, committee member p^* (as in the FA1 algorithm above) could be advantageous, the concrete probabilistic phenomenon arising by this procedure involves two competing effects and deserves some further discussion. Continuing at this same level of informality, the standard large-deviation bounds assert that a committee drawn independently from the stake distribution will suffer from a ϵn excess in empirically observed corrupt members with probability no more than

$$\exp(-2\epsilon^2 n)\,. \tag{5}$$

Deterministic assignment of a single committee member as above has two contrary effects: the first effect reduces the number of randomly selected parties to $n - 1$, which is clearly unfavorable in the context of (5); the second effect conditions the probability space from which these remaining members are selected so that the error margin comprises a larger portion of the sample space—this generates a favorable increase in the effective ϵ to $\epsilon n/(n-1)$. The combined effect is favorable, owing to the quadratic dependence on ϵ; specifically, the tail bound (5) is replaced with

$$\begin{aligned} \exp\left(-2\left(\frac{\epsilon n}{n-1}\right)^2 (n-1)\right) &= \exp\left(-2\epsilon^2 n \cdot \left(\frac{n}{n-1}\right)\right) \\ &= \exp\left(-2\epsilon^2 n \cdot \left(1 + \theta(1/n)\right)\right)\,, \end{aligned}$$

which indicates a rather striking linear advantage in terms of the $1/n$ portion of the committee that has been deterministically assigned.

Relevance to practice. As the performance of our schemes critically depends on the considered stake distribution, we evaluate our schemes on real-world stake distributions of major proof-of-stake blockchains Ethereum, Cardano, Solana, and Algorand. More concretely, for each blockchain we consider the distribution that best reflects the "distribution of power" within the consensus protocol, as this is arguably the most suitable candidate distribution for any selection of a committee for some committee-run layer-two protocol within each blockchain's ecosystem.

Explicit evaluation demonstrates that our methods provide orders-of-magnitude improvement in failure probability for all considered blockchains, even for modest committee sizes. As an illustrative example, Figure 1 considers the Ethereum blockchain, and shows the probabilities of electing a committee of size n that contains at least $n/10$ more corrupt members than expected based on the corruption level of the underlying population; for standard independent sampling (IID) and our new schemes FA1 and wFA. For example, in this setting, the failure probabilities of IID for committees of size 100, 500 and 1000 are achieved by FA1 by sizes 44, 164, and 273, respectively. If the character of the committee's task allows for a weighted committee, wFA can provide the same failure probability at committee sizes 22, 50, and 74, respectively.

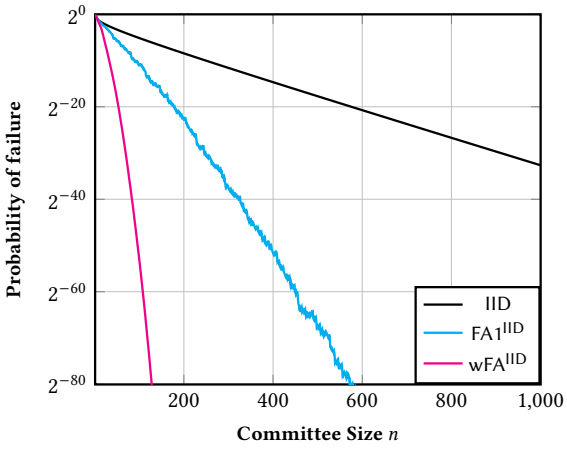

Figure 1: Comparing the probability of electing a committee of size n with ϵ-disproportionate overrepresentation of corrupt parties using the standard independent sampling (IID) and our new schemes FA1$^{\mathsf{IID}}$ and wFA$^{\mathsf{IID}}$ providing a uniform-weight and a weighted committee respectively; for the Ethereum stake distribution and $\epsilon = 0.1$.

Related work. Committee selection is an essential part of the design of any distributed ledger protocol where—for efficiency reasons—consensus is maintained by a selected subset of all participants as opposed to the whole population. Algorand [5] proposes a committee selection method that provides adaptive security. Their approach, based on verifiable random functions [12], has since became standard and, along with the independent stake-based sampling of the committee members, the folklore approach to using a natively permissioned, fixed-population consensus protocol (such as [2–4, 14, 16]) in a permissionless proof-of-stake setting.

Committee selection also appears in the context of sharding. David *et al.* [7] study sharding of a distributed ledger into small shards run by independently selected committees. They consider the standard approach of independently sampling each member of the committee from the underlying population, and observe that the security–efficiency tradeoff leads to impractical committee sizes. Assuming a stake distribution over the underlying population amenable to the *fait accompli* approach, our results would provide a drop-in replacement for the committee-selection part of their scheme with immediate efficiency benefits.

Benhaim *et al.* [1] study the problem of committee selection within the blockchain setting, but focus on using voting schemes for this purpose. These mechanisms allow for smaller committees in their model, as the failure probability vanishes with increasing number of *voters* as opposed to increasing *committee size* in case of lottery-based committee selection considered here. However, this approach is significantly more heavyweight as votes need to be collected and tallied.

Finally, Dimitri [8] considers several variants of the Algorand committee-selection mechanism, but evaluates these on the basis of goals different from ours.

Survey. We formally describe the problem of stake-based committee selection in the next section and give the full version of the fait accompli algorithm in Section 3. In Section 4 we explore the additional power afforded by assigning weights to the selected committee members. We then discuss the relevance to practice in Section 5.

2 STAKE-BASED COMMITTEE SELECTION: FORMAL DESCRIPTION AND STANDARD APPROACHES

We begin by more formally defining the setting described in the introduction.

Notation, conventions, and the basic model. Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Consider a set of parties P along with a staking function $\mathcal{S} : P \to [0, \infty) \subset \mathbb{R}$ describing the stake of each party in P. For convenience, we normalize $\mathcal{S}$ so that it forms a probability distribution—which is to say that $\sum_{p \in P} \mathcal{S}(p) = 1$—and hence refer to $\mathcal{S}$ as the *stake distribution* over P. The general goal is to select a committee from the set P that suitably reflects the stake distribution $\mathcal{S}$.

Committees are treated as multisets, permitting the possibility that they may contain individual parties $p \in P$ multiple times; such parties are then weighted accordingly in terms of the final determination of the fraction of corrupt committee members. To reflect this notationally, we treat a committee of size n as an (ordered) sequence $(c_1, \ldots, c_n)$; while the ordering present in this notation is irrelevant—that is, any permutation of this vector determines the same committee—it's convenient for various aspects of our presentation and, in particular, we sometimes refer to each position in the committee as a *seat*. For such a committee C we define $|C| = n$.

A *committee-selection scheme* F is a randomized procedure for selecting a committee from P based on $\mathcal{S}$. It takes as input P, $\mathcal{S}$, and the desired committee size n (where typically $n \ll |P|$), and outputs a tuple $C = (c_1, \ldots, c_m) \in P^m$ describing the committee where $\mathbb{E}[|C|] = n$. We also study the more restrictive setting where $m = n$ with certainty, which is to say that the scheme always produces a committee of size exactly n. For any fixed tuple of input data, the output of such a scheme is thus entirely described by a probability distribution on P^*, the set of all tuples of parties.

In practice, a committee-selection scheme F is used in the context of an adversary controlling stake $\alpha < 1/2$ and it is desired to find the smallest committee size n such that a fraction $\alpha+\epsilon$ of committee members are corrupt with probability at most δ. In that setting, n would be a function of $\alpha \in (0, 1/2), \epsilon \in (0, 1-\alpha), \delta \in [0, 1)$ and will depend on $\mathcal{S}$.

Security. We consider an adversary corrupting parties in P in order to gain a disproportionate representation in a committee sampled by F. More concretely, given full knowledge of P, $\mathcal{S}$, n and the scheme F itself, the adversary corrupts a subset of parties $A \subset P$. After that, a committee $C = (c_1, \ldots, c_m) \leftarrow \mathsf{F}(P, \mathcal{S}, n)$ is sampled according to F.

For a committee $C = (c_1, \ldots, c_m)$ and a party $p \in P$, let

$$\#_p(C) = |\{i \in [|C|] \mid c_i = p\}|$$

denote the number of seats in C assigned to p; we overload this notation to apply to sets of parties: in particular

$$\#_A(C) = |\{i \in [|C|] \mid c_i \in A\}|$$

denotes the number of seats in C assigned to parties in $A \subset P$. If $\#_A(C) < (\mathcal{S}(A) + \epsilon) \cdot |C|$, in other words, if the corrupt parties are overrepresented on the committee C by less than an ϵ-fraction of the whole committee, then we say that C is a *good* committee; otherwise the committee is *bad*. The figure of merit—for the procedure F as described above—is the worst-case probability of failure (electing a bad committee), taken over all subsets A. This motivates the following definition.

Definition 2.1 (Security of committee selection). For a party set P, a stake distribution $\mathcal{S} : P \to \mathbb{R}$, a committee-selection scheme F, some $n \in \mathbb{N}$ and $\epsilon > 0$, define

$$\mathsf{E}^{\mathcal{S}}_{n,\epsilon}[\mathsf{F}] = \max_{A \subset P} \Pr_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)} [\#_A(C) \geq (\mathcal{S}(A) + \epsilon) \cdot |C|] \ .$$

We overload the notation to indicate the error generated by the optimal procedure for selecting committees of expected size n:

$$\mathsf{E}^{\mathcal{S}}_{\epsilon}(n) = \inf_{\mathsf{F}} \mathsf{E}^{\mathcal{S}}_{n,\epsilon}[\mathsf{F}] \ .$$

The definition above somewhat simplifies the treatment of parameters outlined earlier in this section. In particular, rather than treating α as a parameter of interest, the figure of merit $\mathsf{E}^{\mathcal{S}}_{\epsilon}$ is simply maximized over all possible choices of α (as it is maximizes over all $A \subset P$). Thus, $\mathsf{E}^{\mathcal{S}}_{\epsilon}$ establishes a bound that applies simultaneously to all values of α. This anticipates the structure of our analysis, where is it convenient to maintain control over all possible adversarial subsets of subsidiary stake distributions that arise naturally in the course of our methods and permits our analysis to focus on a single quantity. We remark that for large n the worst-case α converges to $1/2$; as this is often in the range of interest for practice, this formulation may not substantially interfere with the tightness of our results in these settings. On the other hand, for circumstances that warrant small α, it may be possible to establish stronger bounds by carrying this constraint into the analysis.

2.1 Independent and perfect schemes

We identify some special cases of committee-selection schemes. A scheme F is *fixed-size* if it always samples a committee of the same size, i.e., for all P, $\mathcal{S}$, $n \in \mathbb{N}$ we have $\Pr_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)}[|C| = n] = 1$. A fixed-size committee-selection scheme is called *independent* if, for all P, $\mathcal{S}$, $n \in \mathbb{N}$, the n random variables $(c_1, \ldots, c_n)$ resulting from $\mathsf{F}(P, \mathcal{S}, n)$ are independent. Finally, a committee-selection scheme F is *perfect* if for all P, $\mathcal{S}$, $n \in \mathbb{N}$ and parties $p \in P$ we have $\mathbb{E}_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)}[\#_p(C)] = \mathcal{S}(p) \cdot n$.

We now record a general security bound for all independent and perfect schemes.

Lemma 2.2 (Security of perfect independent schemes). *Let F be a perfect and independent committee-selection scheme. For any party set P, stake distribution $\mathcal{S}$ over P, and $n \in \mathbb{N}$, we have*

$$E^{\mathcal{S}}_{n,\epsilon}[\mathsf{F}] \leq \exp(-2n\epsilon^2) \ .$$

Proof. Fix $(P, \mathcal{S}, n)$ and a subset of corrupt parties $A \subset P$. Let $c_1, \ldots, c_n$ be the committee members chosen independently according to F, i.e., $(c_1, \ldots, c_n) \leftarrow \mathsf{F}(P, \mathcal{S}, n)$. Define X_i to be the indicator random variable for the event that $c_i \in A$; then X_i are independent as F is an independent scheme, and

$$\begin{aligned} &\Pr_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)} [\#_A(C) \geq (\mathcal{S}(A) + \epsilon)n] \\ &\quad = \Pr_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)} \left[\sum_{i \in [n]} X_i \geq (\mathcal{S}(A) + \epsilon)n \right] \\ &\quad = \Pr_{C \leftarrow \mathsf{F}(P,\mathcal{S},n)} \left[\sum_{i \in [n]} X_i - \mathbb{E}[X_i] \geq \epsilon n \right] \leq \exp(-2n\epsilon^2) \ , \end{aligned}$$

where we used that $\sum_{i \in [n]} \mathbb{E}[X_i] = \mathcal{S}(A) \cdot n$ as F is perfect. The last inequality is a direct application of the Hoeffding bound (Theorem A.1 in Appendix A) to the random variables $X_1, \ldots, X_n$ with $\lambda := \epsilon n > 0$. □

Scheme $\mathsf{IID}(P, \mathcal{S}, n)$:

Select committee $C = (c_1, \ldots, c_n)$ as follows:
(1) For each $i \in [n]$, independently sample $c_i \leftarrow \mathcal{S}$.
(2) Return $C = (c_1, \ldots, c_n)$.

Figure 2: The standard committee-selection scheme IID.

2.2 I.I.D. sampling

One classical approach to the committee selection problem is to adopt i.i.d. sampling to determine $C = (c_1, \ldots, c_n)$, where each c_i is drawn independently from the distribution $\mathcal{S}$, as described in Figure 2. We denote this scheme by IID. It has the advantage that it is simple, generic, has performance that is independent of $\mathcal{S}$, it produces a fixed-size committee, and is both perfect and independent, hence Lemma 2.2 applies to it. In fact, we have the following slightly more precise bound.

Lemma 2.3 (Security of i.i.d. sampling). *For any party set P, stake distribution $\mathcal{S}$ over P, and $n \in \mathbb{N}$, we have*

$$E^{\mathcal{S}}_{n,\epsilon}[\mathsf{IID}] \leq \sup_{\theta \in [0,1]} \sum_{k \geq (\theta+\epsilon)n} \mathcal{B}(n, \theta; k) \ , \tag{6}$$

where $\mathcal{B}(n, \mu; k) = \binom{n}{k} \mu^k (1-\mu)^{n-k}$ is the binomial distribution.

Proof. Fix $(P, \mathcal{S}, n)$ and a subset of corrupt parties $A \subset P$. Let $(c_1, \ldots, c_n) \leftarrow \mathsf{IID}(P, \mathcal{S}, n)$ and as before, let X_i be the indicator random variable for the event that $c_i \in A$; then the X_i are independent and $\Pr[X_i = 1] = \mathcal{S}(A)$. Then $\sum_i X_i$ follows the binomial distribution with mean $\mu = \mathcal{S}(A) \cdot n$, hence the event that at least an $(\mathcal{S}(A) + \epsilon)$-fraction of the committee is corrupt has probability $\sum_{k \geq (\mathcal{S}(A)+\epsilon)n} \mathcal{B}(n, \mathcal{S}(A); k)$. We have $\mathcal{S}(A) \in [0, 1]$, this hence justifies the lemma. □

2.3 Local sortition

Another standard approach to committee selection is one that we will call *local sortition* (LS). In this method, each party $p \in P$ is included on the committee independently based on a local biased-coin toss, with the bias reflecting the party's stake in $\mathcal{S}$.

Scheme $\mathsf{LS}(P, \mathcal{S}, n)$:

Select committee $C = (c_1, \ldots, c_{|C|})$ as follows:

(1) For each $p \in P$, sample an independent random variable X_p from a Poisson distribution with rate $\mathcal{S}(p) \cdot n$.
(2) Return arbitrary C satisfying $\#_p(C) = X_p$ for all $p \in P$.

Figure 3: The standard committee-selection scheme LS.

In proof-of-stake protocols such as [5, 6], local sortition is often deployed with further features that are orthogonal to our discussion here: the sampling of the local random variables is typically implemented via verifiable random functions [12] so as to provide both privacy and public verifiability of the sortition.

For convenience, we model local sortition as a Poisson process, and describe it in Fig. 3. Notice that the sortition procedure of [5] (as described in [9, Section 5.1]) converges to our Poisson-based description as the total stake in the system goes to infinity.

It is easy to see that local sortition is not a fixed-size committee-selection scheme, yet it is perfect.

Lemma 2.4 (Security of local sortition). *For any party set P, stake distribution $\mathcal{S}$ over P, and $n \in \mathbb{N}$, we have*

$$E^{\mathcal{S}}_{n,\epsilon}[\mathsf{LS}] \le \left(\frac{1}{e}\left(\frac{e}{1+\epsilon}\right)^{(1+\epsilon)}\right)^n + \left(\frac{1}{e}\left(\frac{e}{1-\epsilon}\right)^{(1-\epsilon)}\right)^n . \qquad (7)$$

Proof. Observe that it suffices that $\#_A(C) \le (\mathcal{S}(A) + \epsilon)n$ and for $\#_H(C) \ge ((1-\mathcal{S}(A)) - \epsilon)n$, where we use the notation $\#_H(C) = |C| - \#_A(C)$ (equal to $\#_{\overline{A}}(C)$) to denote the number of "honest" committee members, as the ratio $\#_A(C)/|C|$ is then no more than

$$\frac{(\mathcal{S}(A) + \epsilon)n}{[(1 - \mathcal{S}(A)) - \epsilon]n + [\mathcal{S}(A) + \epsilon]n} = \mathcal{S}(A) + \epsilon .$$

Let P_A denote the Poisson distributed random variable reflecting the total number of committee members in A and, likewise, define P_H to denote the Poisson distributed number of committee members in $\overline{A}$. Defining $\alpha = \mathcal{S}(A)$, note that $\mathbb{E}[P_A] = \alpha n$, $\mathbb{E}[P_H] = (1-\alpha)n$, and that $\alpha \le 1$ by assumption.

In light of the bounds for the tails of the Poisson distribution recorded in Appendix A, we note that

$$\begin{aligned}\Pr[P_A \ge (\alpha+\epsilon)n] &\le \left(\frac{e\alpha}{\alpha+\epsilon}\right)^{(\alpha+\epsilon)n} e^{-\alpha n} \\ &= \left(\left(\frac{e\alpha}{\alpha+\epsilon}\right)^{(\alpha+\epsilon)} e^{-\alpha}\right)^n , \\ &\le \left(\left(\frac{e}{1+\epsilon}\right)^{(1+\epsilon)} e^{-1}\right)^n ,\end{aligned}$$

as this tail is maximized when $\alpha = 1$. A similar calculation yields

$$\begin{aligned}\Pr[P_H \le (1-\alpha-\epsilon)n] &\le \left(\frac{e(1-\alpha)n}{(1-\alpha-\epsilon)n}\right)^{(1-\alpha-\epsilon)n} e^{-(1-\alpha)n} \\ &\le \left(\left(\frac{e}{1-\epsilon}\right)^{(1-\epsilon)} e^{-1}\right)^n\end{aligned}$$

as this tail is maximized when $\alpha = 0$. The probability that either of these events occurs is no more than the sum of the probabilities, which yields the statement of the theorem. □

3 UNIFORM-WEIGHT FAIT ACCOMPLI COMMITTEES

We begin with a detailed presentation and analysis of the scheme discussed in the introduction. While we will later formulate schemes with improved performance, several technical features of the analysis will be useful during the subsequent development. Furthermore, we also present a local-sortition variant of this scheme which is of particular interest when we consider committee selection in the setting of an adaptive adversary.

In general, the performance of the schemes we study in this section are reflected by the ability of functions taking values in the set $\frac{1}{n}\mathbb{N}_0 = \{0, 1/n, 2/n, \ldots\}$ to accurately approximate the stake distribution. Two different, but closely related, notions shall in fact be of interest.

Definition 3.1. Let P be a finite set. We say that a function $f : P \to \mathbb{R}$ is *n-integral* if $f(p) \in \frac{1}{n}\mathbb{N}_0 = \{0, 1/n, 2/n, \ldots\}$ for each $p \in P$. For a probability distribution $\mathcal{S} : P \to \mathbb{R}$ we define

$$\begin{aligned}\tau_1(\mathcal{S}; n) &= \min_{\substack{f: P \to \frac{1}{n}\mathbb{N}_0 \\ \|f\| \le 1}} \|\mathcal{S} - f\| , \text{ and} \\ \tau_{\mathfrak{b}}(\mathcal{S}; n) &= \min_{\substack{f: P \to \frac{1}{n}\mathbb{N}_0 \\ f \le \mathcal{S}}} \|\mathcal{S} - f\| ,\end{aligned}$$

where the notation $f \le \mathcal{S}$ means that $f(p) \le \mathcal{S}(p)$ for each $p \in P$, and $\|f\|$ denotes the 1-norm: $\|f\| = \sum_{p \in P} |f(p)|$.

Witnesses. In the context of a particular function $\mathcal{S} : P \to \mathbb{R}$, we say that $f : P \to \frac{1}{n}\mathbb{N}_0$ *witnesses* τ_1 if it achieves the minimum that defines τ_1. We use this same terminology for $\tau_{\mathfrak{b}}$ (with the understanding that f meets the criteria for $\tau_{\mathfrak{b}}$: $f \le \mathcal{S}$).

Note that for any distribution $\mathcal{S}$ we have

$$\tau_1(\mathcal{S}; n) \le \tau_{\mathfrak{b}}(\mathcal{S}; n)$$

as the set of functions over which τ_1 minimizes contains that of $\tau_{\mathfrak{b}}$.

3.1 Scheme FA1

Returning to the scheme presented in the introduction, note that the scheme has two logical "phases," the first in which deterministic assignments are made for parties for which $\mathcal{S}(p) \ge 1/n$, and the second "fallback phase" in which stochastic assignments for the rest of the committee are made. We note that the fallback phase can in fact be instantiated with either of the two conventional committee selection schemes defined in the last section. A detailed description of this committee-selection scheme FA1 is given in Figure 4, parametrized by a fallback scheme F which, by itself, must also be a committee-selection procedure. In a nutshell, FA1 provides benefits over the standard methods for any stake distribution that contains parties controlling at least a $1/n$-fraction of stake; if no such parties exist, $\mathsf{FA1}^{\mathsf{F}}$ invokes the fallback scheme F.

Theorem 3.2 (Security of $\mathsf{FA1}^{\mathsf{IID}}$). *Let P be a party set, $\mathcal{S}$ be a stake distribution over P, and $n \in \mathbb{N}$. Let $\tau = \tau_{\mathfrak{b}}(\mathcal{S}, n)$. If $\tau = 0$ then*

$$E^{\mathcal{S}}_{\epsilon}(n) = E^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{IID}}\right] = 0 ;$$

Scheme $\mathsf{FA1}^{\mathsf{F}}(P, \mathcal{S}, n)$:

Select committee $C = (c_1, \ldots, c_m)$ as follows:

(1) Let $\mathcal{S}_\flat : P \to \frac{1}{n}\mathbb{N}_0$ witness $\tau_\flat(\mathcal{S}; n)$. In particular
$$\mathcal{S}_\flat(p) = \frac{\lfloor \mathcal{S}(p)n \rfloor}{n} .$$
For each $p \in P$, assign $\mathcal{S}_\flat(p) \cdot n$ seats to p. Let $(c_1, c_2, \ldots, c_{n_1})$ be the $n_1 \triangleq \|\mathcal{S}_\flat\| \cdot n \le n$ seats assigned in this way.

(2) If $n_1 < n$, define a distribution $\mathcal{S}_1 : P \to [0, 1]$ as
$$\mathcal{S}_1(p) = \left(\mathcal{S}(p) - \mathcal{S}_\flat(p)\right) / \tau_\flat(\mathcal{S}; n)$$
and let $(c_{n_1+1}, \ldots, c_{n_1+n_2}) \leftarrow \mathsf{F}(P, \mathcal{S}_1, n - n_1)$; otherwise set $n_2 = 0$.

(3) Return $C = (c_1, \ldots, c_m)$ with $m = n_1 + n_2$.

Figure 4: Committee-selection scheme FA1 parametrized by a fallback scheme F.

otherwise

$$E^{\mathcal{S}}_{\epsilon}(n) \le E^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{IID}}\right] \le \exp(-2n\epsilon^2/\tau) .$$

This follows from the following lemma.

Lemma 3.3. *Let P be a party set, $\mathcal{S}$ be a stake distribution over P, and $n \in \mathbb{N}$. Define*
$$\mathcal{S}_\flat(p) = \frac{\lfloor n\mathcal{S}(p) \rfloor}{n}$$
and $\tau = 1 - \sum_{p \in P} \mathcal{S}_\flat(p)$. Then $\tau = \tau_\flat(\mathcal{S}; n)$ and, moreover, if $\tau > 0$ then for any fixed-size committee-selection scheme F, we have
$$E^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{F}}\right] \le E^{(\mathcal{S}-\mathcal{S}_\flat)/\tau}_{\tau n, \epsilon/\tau}[\mathsf{F}] . \tag{8}$$

Proof. Fix some $(P, \mathcal{S}, n)$ and observe that
$$\mathcal{S}(p) - 1/n \le \mathcal{S}_\flat(p) \le \mathcal{S}(p)$$
for all p and that $\mathcal{S}_\flat$ takes values in $\frac{1}{n}\mathbb{N}_0 \triangleq \{k/n \mid k \in \mathbb{N}_0\}$, hence
$$\mathcal{S}_\flat = \operatorname*{arg\,max}_{\substack{f : P \to \frac{1}{n}\mathbb{N}_0 \\ f \le \mathcal{S}}} \sum_{p \in P} f(p) ,$$
establishing $\tau = \tau_\flat(\mathcal{S}; n)$; below we assume $\tau > 0$.

Let F be any fixed-size committee-selection scheme. Let $\mathcal{S}_1$ denote the distribution $(\mathcal{S} - \mathcal{S}_\flat)/\tau$, and let $\tilde{C}$ and C^* denote the output distributions of $\mathsf{FA1}^{\mathsf{F}}(P, \mathcal{S}, n)$ and $\mathsf{F}(P, \mathcal{S}_1, \tau n)$, respectively. Let $\tilde{A} \subset P$ be the set witnessing $\mathsf{E}^{\mathcal{S}}_{n,\epsilon}[\mathsf{FA1}^{\mathsf{F}}]$, i.e., the one maximizing the probability $\Pr_{C \leftarrow \tilde{C}}[\#_{\tilde{A}}(C) \ge (\mathcal{S}(\tilde{A}) + \epsilon) \cdot n]$.

First, observe that we can without loss of generality assume that each party $p \in P$ satisfies either $\mathcal{S}(p) \in \frac{1}{n}\mathbb{N} = \{1/n, 2/n, \ldots\}$ or $\mathcal{S}(p) < 1/n$. This is because any party that does not satisfy any of these two conditions can be "split" into two virtual parties p_1 and p_2 such that $\mathcal{S}(p_1) = \mathcal{S}_\flat(p)$ and $\mathcal{S}(p_2) = \mathcal{S}(p) - \mathcal{S}_\flat(p)$, resulting in two parties that both satisfy the above condition. Note that this replacement can only make the adversary stronger, as the two virtual parties can now be corrupted independently.

Given the above assumption, let $P_1 \triangleq \left\{p \in P \mid \mathcal{S}(p) \in \frac{1}{n}\mathbb{N}\right\}$ and $P_2 \triangleq \{p \in P \mid \mathcal{S}(p) < 1/n\}$, then P can be partitioned as $P = P_1 \cup P_2$. Denote $\tilde{A}_i = \tilde{A} \cap P_i$ for both $i \in \{1, 2\}$.

We first observe that for any $p_i \in P_1$, according to step (1) of FA1, p_i will be assigned exactly $\mathcal{S}_\flat(p_i) \cdot n$ committee seats. Therefore, for any $C \leftarrow \tilde{C}$ we have
$$\#_{\tilde{A}_1}(C) = \mathcal{S}(\tilde{A}_1) \cdot n . \tag{9}$$

On the other hand, by definition of $\mathcal{S}_1$, for any $p \in P_2$ we have $\mathcal{S}_1(p) = \mathcal{S}(p)/\tau$, and hence
$$\mathcal{S}_1(\tilde{A}_2) = \mathcal{S}(\tilde{A}_2)/\tau . \tag{10}$$

Using these observations, we have
$$\begin{aligned}
\mathsf{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{F}}\right] &\overset{(a)}{=} \Pr_{C \leftarrow \tilde{C}}\left[\#_{\tilde{A}}(C) \ge (\mathcal{S}(\tilde{A}) + \epsilon) \cdot n\right] \\
&\overset{(b)}{=} \Pr_{C \leftarrow \tilde{C}}\left[\#_{\tilde{A}_2}(C) \ge (\mathcal{S}(\tilde{A}_2) + \epsilon) \cdot n\right] \\
&\overset{(c)}{=} \Pr_{C' \leftarrow C^*}\left[\#_{\tilde{A}_2}(C') \ge (\mathcal{S}_1(\tilde{A}_2)\tau + \epsilon) \cdot n\right] \\
&= \Pr_{C' \leftarrow C^*}\left[\#_{\tilde{A}_2}(C') \ge (\mathcal{S}_1(\tilde{A}_2) + \epsilon/\tau) \cdot \tau n\right] \\
&\le \mathsf{E}^{\mathcal{S}_1}_{\tau n, \epsilon/\tau}[\mathsf{F}] ,
\end{aligned}$$
where (a) follows from the definition of $\tilde{A}$; (b) uses (9) and the fact that $\tilde{A}$ can be partitioned as $\tilde{A}_1 \cup \tilde{A}_2$; and (c) follows from (10). □

Theorem 3.2 can now be easily obtained by applying Lemma 3.3 to $\mathsf{FA1}^{\mathsf{IID}}$ and subsequently invoking Lemmas 2.2 for IID. Moreover, it is easy to observe that if $\tau_\flat(\mathcal{S}, n) = 0$ then $\mathsf{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{F}}\right] = 0$ for any F and $\epsilon > 0$; in particular if F is a fixed-size scheme, then $\mathsf{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{F}}\right] = 0$ holds for any $\tau_\flat < \epsilon$.

We now analyze an adaptively secure variant of FA1, obtained by using LS as the fallback scheme.

Theorem 3.4 (Security of $\mathsf{FA1}^{\mathsf{LS}}$). *Let P be a party set, $\mathcal{S}$ be a stake distribution over P, and $n \in \mathbb{N}$. Let $\tau = \tau_\flat(\mathcal{S}, n)$. Then*
$$E^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{FA1}^{\mathsf{LS}}\right] \le \left(\frac{1}{e^\tau}\left(\frac{e\tau}{\tau+\epsilon}\right)^{(\tau+\epsilon)}\right)^n + \left(\frac{1}{e^\tau}\left(\frac{e\tau}{\tau-\epsilon}\right)^{(\tau-\epsilon)}\right)^n \tag{11}$$
unless $\tau \le \epsilon$, in which case the second term should be treated as zero.

Proof. As in the proof of Lemma 2.4, we note that it suffices for $\#_A(C) \le (\mathcal{S}(A) + \epsilon)n$ and for $\#_H(C) \ge ((1 - \mathcal{S}(A)) - \epsilon)n$, referring to the final committee C chosen by the scheme. (We again use the notation $\#_H(C) = |C| - \#_A(C)$.) After the deterministic assignment of committee members is complete—determining an initial committee C_1—the local sortition scheme LS is called upon to draw a committee C_2 with expected size τn; C is then the union of the two committees. As the scheme guarantees that
$$\mathbb{E}[\#_A(C_1) + \#_A(C_2)] = \mathcal{S}(A)n$$
and C_1 is deterministic,
$$\mathbb{E}[\#_A(C_2)] = \mathcal{S}(A)n - \#_A(C_1)$$
and we define $\alpha n = \mathbb{E}[\#_A(C_2)]$; clearly $\alpha n \le \tau n$, as τn is the expected size of C_2.

As $|C_1|$ is fixed, any deviation in the statistics $\#_A(C)$ and $\#_H(C)$ from their expected values are determined by C_2, which is to say that

$$\#_A(C) \geq \mathbb{E}[\#_A(C)] + \ell \quad \Leftrightarrow \quad \#_A(C_2) \geq \mathbb{E}[\#_A(C_2)] + \ell\,;$$

the same can be said for deviations of $\#_H(C)$. We may thus expand the tail bounds of interest around the Poisson distributed random variables $\#_A(C_2)$ and $\#_H(C_2)$.

In light of the bounds for the tails of the Poisson distribution recorded in Appendix A, we note that

$$\begin{aligned}\Pr[\#_A(C_2)] \geq (\alpha+\epsilon)n] &\leq \left(\frac{e\alpha}{\alpha+\epsilon}\right)^{(\alpha+\epsilon)n} e^{-\alpha n} \\ &\leq \left(\left(\frac{e\tau}{\tau+\epsilon}\right)^{(\tau+\epsilon)} e^{-\tau}\right)^n,\end{aligned}$$

as this tail is maximized when $\alpha = \tau$. A similar calculation yields

$$\begin{aligned}\Pr[\#_H(C_2) \leq (\tau-\alpha-\epsilon)n] &\leq \left(\frac{e(\tau-\alpha)n}{(\tau-\alpha-\epsilon)n}\right)^{(\tau-\alpha-\epsilon)n} e^{-(\tau-\alpha)n} \\ &\leq \left(\left(\frac{e\tau}{\tau-\epsilon}\right)^{(\tau-\epsilon)} e^{-\tau}\right)^n\end{aligned}$$

as this tail is maximized when $\alpha = 0$. Observe that the probability of this tail is zero (for any α) when $\tau < \epsilon$. The probability that either of these events occurs is no more than the sum of the probabilities, which yields the statement of the theorem. □

3.2 Scheme FA2

We now describe a more refined algorithm FA2, given in Figure 5. While the algorithm asymptotically improves on FA1, demonstrating that security can scale with τ_1 (rather than the larger value $\tau_\flat$), we are not aware of sufficiently strong large deviation bounds parametrized by variance to yield effective improvements in practice. For this reason, we primarily include this for theoretical interest.

Theorem 3.5 (Security of FA2). *Let $\mathcal{S} : P \to \mathbb{R}$ be a probability distribution. Let $\tau_1 = \tau_1(S; n)$ and define $\sigma^2 = 3\tau_1/4$. If $\tau_1 = 0$ then $E^{\mathcal{S}}_{n,\epsilon}[\mathsf{FA2}] = 0$, otherwise*

$$E^{\mathcal{S}}_{n,\epsilon}[\mathsf{FA2}] \leq \left(\left(1+\frac{\epsilon}{\sigma^2}\right)^{-\frac{\epsilon+\sigma^2}{1+\sigma^2}} (1-\epsilon)^{-\frac{1-\epsilon}{1+\sigma^2}}\right)^n. \tag{12}$$

Proof. As in the description of the scheme, let f be an n-integral function witnessing $\tau_1(\mathcal{S}, n)$ and define, for each $p \in P$,

$$\mathcal{S}_\flat(p) = \lfloor n\mathcal{S}(p)\rfloor/n \qquad \text{and} \qquad \mathcal{S}^\sharp(p) = \lceil n\mathcal{S}(p)\rceil/n\,.$$

Then $\mathcal{S}_\flat \leq \mathcal{S} \leq \mathcal{S}^\sharp$ and, as $\|\mathcal{S}_\flat\| \leq \|\mathcal{S}\| \leq 1$, it's easy to see that any witness function f for τ_1 satisfies $\mathcal{S}_\flat \leq f$ and, in fact, $\forall p, f(p) \in \{\mathcal{S}_\flat(p), \mathcal{S}^\sharp(p)\}$. We then write $\mathcal{S} = f + \alpha - \beta$ where $\alpha, \beta : P \to \mathbb{R}$ are non-negative functions with disjoint supports (which is to say that they are never both positive). Then the support of β is the set $P_f = \{p \in P \mid f(p) > \mathcal{S}(p)\}$ and, if β is nonzero on p, it takes the value $\mathcal{S}^\sharp(p) - \mathcal{S}(p)$; similarly, if α is nonzero on p, it takes the value $\mathcal{S}(p) - \mathcal{S}_\flat(p)$. Finally, we note that

$$\tau_1 = \|\mathcal{S} - f\| = \|\alpha\| + \|\beta\|\,.$$

Scheme FA2$(P, \mathcal{S}, n)$:

Select committee $C = (c_1, \ldots, c_n)$ as follows:

(1) Determine a function f witnessing $\tau_1(\mathcal{S}; n)$, i.e., such that
$$f = \operatorname*{arg\,min}_{\substack{f:\, P \to \frac{1}{n}\mathbb{N}_0,\\ \|f\| \leq 1}} \|\mathcal{S} - f\|\,,$$
and let
$$P_f = \{p \in P \mid f(p) > \mathcal{S}(p)\}\,.$$
Define, also, $\mathcal{S}_\flat(p) = \lfloor n\mathcal{S}(p)\rfloor/n$ and $n_1 = n \cdot \|\mathcal{S}_\flat\|$.

(2) Assign n_1 seats $(c_1, \ldots, c_{n_1})$ as in FA1. That is, if $n\mathcal{S}_\flat(p) = k$, assign k of these seats to p.

(3) For each $p \in P_f$, assign a seat to p with probability $1 - (f(p) - \mathcal{S}(p)) \cdot n$, otherwise leave it unassigned. Let $(c_{n_1+1}, c_{n_1+2}, \ldots, c_{n_1+n_2})$ be the $n_2 \leq |P_f|$ seats assigned in this way.

(4) If $n_1 + n_2 < n$, let $r = \sum_{p \notin P_f}(\mathcal{S}(p) - f(p))$, define a distribution $\mathcal{S}_2 : P \to [0,1]$ as
$$\mathcal{S}_2(p) = \begin{cases} 0 & \text{if } p \in P_f, \\ (\mathcal{S}(p) - f(p))/r & \text{if } p \notin P_f, \end{cases}$$
and let $(c_{n_1+n_2+1}, \ldots, c_{n_1+n_2+n_3}) \leftarrow \mathsf{IID}(P, \mathcal{S}_2, n - n_1 - n_2)$, otherwise set $n_3 = 0$.

(5) Return $C = (c_1, \ldots, c_n)$; note $n = n_1 + n_2 + n_3$.

Figure 5: Committee-selection scheme FA2.

To prepare for the remainder of the analysis, we identify the (committee seat) random variables defined in the scheme and establish that the scheme is indeed perfect.

- The scheme calls (in step (2)) for $n_1 = \|\mathcal{S}_\flat\|n$ seats to be assigned deterministically according to $\mathcal{S}_\flat$ (so that the number of seats assigned to p is exactly $\mathcal{S}_\flat(p)n$). Fixing a particular ordering for these and treating these deterministic choices as probability distributions with a single nonzero value, let $C_1, \ldots, C_{n_1}$ denote the associated probability distributions (on P) for these first n_1 seats. We remark that $(1/n)\sum_i^{n_1} C_i = \mathcal{S}_\flat$.
- The scheme then assigns $|P_f|$ additional seats, out of which n_2 are assigned in step (3) and the rest is filled in step (4). We begin by verifying that $n_1 + |P_f| \leq n$. Noting that $f = \mathcal{S}^\sharp$ for all $p \in P_f$, it follows that $\|f\| \geq \|\mathcal{S}_\flat\| + (1/n) \cdot |P_f|$ and hence that $n \geq n\|f\| \geq n\|\mathcal{S}_\flat\| + |P_f| = n_1 + |P_f|$.
  Recall that for each $p \in P_f$, a committee member is independently assigned according to the probability distribution C_p defined so that p is selected with probability
  $$1 - \beta(p) \cdot n = 1 - n(\mathcal{S}^\sharp(p) - \mathcal{S}(p)) = n(\mathcal{S}(p) - \mathcal{S}_\flat(p))$$
  and, otherwise, is drawn from the distribution $\alpha/\|\alpha\|$ supported on $\overline{P_f}$. As a matter of bookkeeping, we write $P_f = \{p_1, \ldots, p_{|P_f|}\}$ and, for each $i \in \{1, \ldots, |P_f|\}$, define C_{n_1+i} to be C_{p_i}.

We remark that for any $p \in P_f$,

$$\frac{1}{n}\sum_{i=1}^{n_1+|P_f|} C_i(p) = \mathcal{S}_\flat(p) + \frac{1}{n}\sum_{i=1}^{|P_f|} C_{n_1+i}(p)$$
$$= \mathcal{S}_\flat(p) + (\mathcal{S}(p) - \mathcal{S}_\flat(p)) = \mathcal{S}(p)\,,$$

as p has nonzero contribution from exactly one C_i with $n_1 < i \le n_1 + |P_f|$. Restricted to the set $\overline{P_f}$, each C_{p_i} is proportional to α and hence the sum $(1/n)\sum_i C_i(p)$ above is $\mathcal{S}_\flat + c\alpha$, for a scalar c.

- Finally, the remaining members are chosen from the distribution $\alpha/\|\alpha\|$ in step (4). Continuing the naming convention, we define $C_i = \alpha/\|\alpha\|$ for $n_1 + |P_f| < i \le n$. In light of the running calculations above (of the sums $\sum_i C_i$), we conclude that $(1/n)\cdot\sum_i C_i = \mathcal{S}$, as desired; the scheme is perfect.

Let c_i (for $0 < i \le n$) denote independent random variables taking values in P, each distributed according to C_i. For a particular set $A \subset P$ of parties (corrupted by the adversary) let 1_A denote the characteristic function for A and consider the random variables

$$X_i = 1_A(c_i) = \begin{cases} 1 & \text{if } c_i \text{ is corrupt,} \\ 0 & \text{otherwise.} \end{cases}$$

We let $\mathcal{S}(A)$ denote $\sum_{a\in A}\mathcal{S}(a)$ and note that that $\sum \mathbb{E}[X_i] = \mathcal{S}(A)$ as the scheme is perfect.

To complete the proof, we apply a large deviation bound for independent random variables parametrized by variance due to Hoeffding [10]. (A detailed description of the inequality appears in Appendix A.) We recall that for a $\{0,1\}$-valued random variable Z, the variance $\mathbb{V}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$ is equal to $\mathbb{E}[Z](1 - \mathbb{E}[Z])$. We consider the variance of the random variables X_i:

- $\mathbb{V}[X_1] = \cdots = \mathbb{V}[X_{n_1}] = 0$, as these random variables are constants.
- Recalling that $X_{n_1+i} = 1_A(p_i)$ with probability $1 - \beta(p_i)n$, we see that $\mathbb{V}[X_{n_1+i}] \le \beta(p_i)n \cdot (1 - \beta(p_i)n) \le \beta(p_i)n$ (for $0 < i \le |P_f|$). Here we continue to use the notation above: $P_f = \{p_1, \ldots, p_{|P_f|}\}$.
- Finally, for the remaining random variables $X_{n_1+|P_f|+i}$ we invoke the trivial bound, true for all $\{0,1\}$-valued random variables: $\mathbb{V}[X_i] \le 1/4$. Observe that the total probability mass of the $|P_f| - n_2$ variables selected in the second phase that is not assigned to the "target" p_i but rather drawn from $\alpha/\|\alpha\|$ is exactly $n\|\beta\|$. It follows that there are $n(\|\alpha\| - \|\beta\|)$ random variables of this third type.

The total variance is thus bounded by

$$\sum_{p\in P_f} n\beta(p) + \frac{n}{4}(\|\alpha\| - \|\beta\|) \le n\|\beta\| + \frac{n}{4}(\|\alpha\| - \|\beta\|)$$
$$\le \frac{3n}{4}(\|\alpha\| + \|\beta\|) = \frac{3n\tau_1}{4}\,.$$

Finally, we apply the Hoeffding inequality mentioned above (Theorem A.3 of Appendix A). Define $Z_i = X_i - \mathbb{E}[X_i]$ and note that $\mathbb{E}[Z_i] = 0$, $\mathbb{E}[Z_i^2] = \mathbb{V}[X_i]$ and that $\sum_i Z_i - \mathbb{E}[\sum_i Z_i] = \sum_i Z_i = \sum_i X_i - \mathbb{E}[\sum_i X_i]$. Applying the inequality to the Z_i yields

$$\Pr\left[\sum_i X_i - n\mathcal{S}(A) \ge \epsilon n\right] \le \left(\left(1 + \frac{\epsilon}{v}\right)^{-\frac{\epsilon+v}{1+v}}(1-\epsilon)^{-\frac{1-\epsilon}{1+v}}\right)^n$$

for $v := 3\tau_1/4$, as desired. □

4 WEIGHTED COMMITTEES

In this section we consider an extension of the basic setting of Section 2 in which we allow committees to be *weighted*: intuitively, a committee $C = (c_1, \ldots, c_{|C|})$ is now a set of parties with associated weights that sum up to 1; the weight $\mathrm{w}_C(i)$ of the committee member c_i describes her "voting power" in the committee. In other words, assigning a seat c_i with weight $\mathrm{w}_C(i)$ to a party p is meant to have the same effect as assigning a $(\mathrm{w}_C(i)/\sum_j \mathrm{w}_C(j))$-fraction of all committee seats to this party in the uniform setting of Section 2.

4.1 Model

Notation. As before, we consider a set of parties P and a stake distribution $\mathcal{S} : P \to [0,\infty)$ satisfying $\sum_{p\in P}\mathcal{S}(p) = 1$. A weighted committee is a pair (C, w_C), where $C \subset P$ is the set of committee members, and $\mathrm{w}_C\colon [|C|] \to [0,1]$ is a weight function assigning a weight to each seat of the committee. We again permit randomized committee-selection procedures that produce variable-sized committees, as long as the expected committee size can be controlled; similarly, for convenience we also ask that the expected sum of weights in a committee is 1.

More formally, a *weighted committee-selection scheme* W is a randomized procedure for selecting a committee from P based on $\mathcal{S}$. It takes as input P, $\mathcal{S}$, and n, and outputs a tuple (C, w_C) where $C = (c_1, \ldots, c_{|C|}) \in P^{|C|}$ satisfies $\mathbb{E}[|C|] = n$; and $\mathrm{w}_C\colon [|C|] \to [0,1]$ satisfies $\mathbb{E}\left[\sum_{i\in[|C|]}\mathrm{w}_C(i)\right] = 1$. The adversary can again corrupt an arbitrary subset $A \subset P$ of parties, with full knowledge of the above procedure. Finally, a weighted committee $(C, \mathrm{w}_C) \leftarrow \mathsf{W}(P, \mathcal{S}, n)$ is sampled according to W.

Security. The definition below is an analogue of Def. 2.1 for the weighted-committee case. For convenience, we overload the notation to also apply $\mathrm{w}_C(\cdot)$ to sets of parties $B \subseteq P$, so that

$$\mathrm{w}_C(B) = \sum_{\substack{i\in[|C|] \\ c_i\in B}} \mathrm{w}_C(i)\,,$$

hence $\mathrm{w}_C(P)$ denotes the total sum of weights in the committee.

Definition 4.1 (Security; the weighted case). For party set P, a stake distribution $\mathcal{S} : P \to \mathbb{R}$, a weighted committee-selection scheme W, some $n \in \mathbb{N}$ and $\epsilon > 0$, define

$$\overline{\mathrm{E}}^{\mathcal{S}}_{n,\epsilon}[\mathsf{W}] = \max_{A\subset P}\Pr_{(C,\mathrm{w}_C)\leftarrow\mathsf{W}(P,\mathcal{S},n)}\left[\mathrm{w}_C(A) \ge (\mathcal{S}(A) + \epsilon)\cdot\mathrm{w}_C(P)\right]\,,$$

where $C = (c_1, \ldots, c_{|C|})$. We again overload the notation by defining

$$\overline{\mathrm{E}}^{\mathcal{S}}_{\epsilon}(n) = \inf_{\mathsf{W}}\overline{\mathrm{E}}^{\mathcal{S}}_{n,\epsilon}[\mathsf{W}]\,.$$

Observe that the uniform-weight setting defined in Section 2 is in fact a special case of the weighted setting. In this section, we will refer to schemes of the special type defined in Section 2 as *uniform-weight schemes*.

4.2 Scheme wFA

We now present our committee-selection scheme for the weighted setting, denoted wFA, which is given in Figure 6.

Scheme $\mathsf{wFA}^{\mathsf{F}}(P, \mathcal{S}, n)$:

Select weighted committee (C, w_C) with $C = (c_1, \ldots, c_{|C|})$ as follows:

(1) Let $(p_1, \ldots, p_{|P|})$ be an ordering of the set P that is non-increasing with respect to stake $\mathcal{S}(\cdot)$. For each $i \in [n-1]$, let $\rho_i = \sum_{j=i}^{|P|} \mathcal{S}(p_j)$. Let i^* be the smallest $i \in \mathbb{N}$ such that either $\rho_i = 0$ or

$$\left(1 - \frac{\mathcal{S}(p_i)}{\rho_i}\right)^2 \geq \frac{n-i}{n-i+1}\,, \tag{13}$$

note that $i^* \leq n$. For each $i \in [i^* - 1]$, assign the seat c_i to party p_i (i.e., set $c_i = p_i$) and set $\mathsf{w}_C(i) = \mathcal{S}(p_i)$. Let $(c_1, c_2, \ldots, c_{n_1})$ be the $n_1 = i^* - 1 < n$ seats assigned in this way.

(2) Denote

$$\rho \triangleq \rho_{i^*} \tag{14}$$

and if $\rho > 0$, define a distribution $\mathcal{S}_3 : P \to [0, 1]$ as

$$\mathcal{S}_3(p_i) = \begin{cases} 0 & \text{if } i \leq n_1, \\ \mathcal{S}(p_i)/\rho & \text{otherwise,} \end{cases} \tag{15}$$

and let $(c_{n_1+1}, \ldots, c_{n_1+n_2}) \leftarrow \mathsf{F}(P, \mathcal{S}_3, n - n_1)$. For all $i \in \{n_1 + 1, \ldots, n_1 + n_2\}$, set $\mathsf{w}_C(i) = \rho/n_2$.

(3) Return (C, w_C) where $C = (c_1, \ldots, c_{|C|})$ with $|C| = n_1 + n_2$.

Figure 6: Weighted committee-selection scheme wFA **parametrized by a fallback (uniform-weight) scheme** F**.**

The scheme is, as before, parametrized by a "fallback" uniform-weight committee-selection scheme F. In essence, wFA proceeds as follows: it considers individual parties $p \in P$ in a non-increasing order with respect to their stake $\mathcal{S}(p)$ by gradually picking up the largest-stake yet-unprocessed party p_i and assigning it a seat c_i on the weighted committee, with a weight $\mathcal{S}(p_i)$ equal to its stake in the stake distribution $\mathcal{S}$. This process continues as long as the stake of p_i is sufficient to not trigger a particular stopping condition expressed in (13). After the stopping condition is triggered, the remaining seats on the committee are assigned using the fallback scheme F, attributing uniform weights to the committee members assigned in this second step.

Let us illustrate the gains provided by $\mathsf{wFA}^{\mathsf{IID}}$ compared to both IID and $\mathsf{FA1}^{\mathsf{IID}}$ using the same toy example we considered in the introduction. Continuing with that scenario, if wFA is used to choose a committee of size $n = 30$, it would assign the first seat in that committee to the party p_{big} with weight $\mathsf{w}_C(1) = \mathcal{S}(p_{\text{big}}) = 1/3$, and the remaining 29 seats would be assigned using IID to the minor stakeholders with the remaining weight distributed evenly, i.e., each of the 29 committee seats would come with weight $(1-1/3)/29 = 2/87$. Just like in the case of FA1 discussed in the introduction, it would be counterproductive for the adversary to corrupt p_{big}, hence his best strategy is to corrupt $1/2$ of all the minor stakeholders, amounting to $1/3$ of total stake. Then the probability that the adversary achieves his goal is

$$\Pr\left[B(29, 1/2) \geq 22\right] \approx 0.4\%\,,$$

as this requires that at least 22 out of the 29 seats assigned by IID would end up controlled by a corrupted party—since, recall, the weight of each of these seats is $2/87$. This error probability compares favorably to the 4% (resp. 2%) adversarial success probability obtained by using IID (resp., $\mathsf{FA1}^{\mathsf{IID}}$). Moreover, one can easily verify that when using wFA, committee size $n = 12$ is sufficient to keep the probability of adversarial success smaller than (4), the error of the IID method obtained for committee size $n = 30$.

The intuition behind the stopping condition (13) is to determine in each iteration whether the currently considered party p_i has enough stake to warrant assigning it a seat on the committee deterministically, as opposed to using a simple scheme such as IID to conclude the committee selection. Namely, if we approximate the error $\mathsf{E}_{n,\epsilon}$ of the IID invoked to assign n seats by $\exp(-2n\epsilon^2)$ (cf. Lemmas 2.2 and 2.3), transitioning to IID in the i-th iteration entails an error of

$$e_i = \exp(-2(n - i + 1)\epsilon^2)\,,$$

while assigning the i-th party p_1 a seat deterministically and then invoking IID in the subsequent iteration $i+1$ would lead to an error

$$e_{i+1} = \exp\left(-2(n-i)\left(\frac{\epsilon}{1 - \mathcal{S}(p_i)/\rho_i}\right)^2\right)\,,$$

as the remaining stake would get re-scaled by the factor $1-\mathcal{S}(p_i)/\rho_i$ after removing the party p_i. Asking whether $e_i \leq e_{i+1}$ translates exactly to the condition (13). Note that the stopping condition will be satisfied at latest when $i = n$, as

$$\left(1 - \frac{\mathcal{S}(p_n)}{\rho_n}\right)^2 \geq 0 = \frac{n-n}{n-n+1}\,,$$

and hence the scheme F will be left to assign at least one seat.

The security of wFA is stated in the following theorem.

Theorem 4.2 (Security of $\mathsf{wFA}^{\mathsf{IID}}$). *Let P be a party set, $\mathcal{S}$ be a stake distribution over P, and $n \in \mathbb{N}$. Let n_1 and ρ be as determined in Figure 6. If $\rho = 0$ then $\overline{E}^{\mathcal{S}}_{\epsilon}(n) = \overline{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{wFA}^{\mathsf{IID}}\right] = 0$; otherwise we have*

$$\overline{E}^{\mathcal{S}}_{\epsilon}(n) \leq \overline{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{wFA}^{\mathsf{IID}}\right] \leq \exp(-2(n - n_1)\epsilon^2/\rho^2)\,.$$

Theorem 4.2 follows easily from the following lemma, which we prove first.

Lemma 4.3. *Let P be a party set, $\mathcal{S}$ be a stake distribution over P, and $n \in \mathbb{N}$. For $n_1, \rho, \mathcal{S}_3$ as in Figure 6 and for any uniform-weight, fixed-size committee-selection scheme* F*, if $\rho > 0$ then we have*

$$\overline{E}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{wFA}^{\mathsf{F}}\right] \leq E^{\mathcal{S}_3}_{n-n_1,\epsilon/\rho}[\mathsf{F}]\,. \tag{16}$$

Proof. Fix some $(P, \mathcal{S}, n)$ and let F be any uniform-weight fixed-size committee-selection scheme; assume $\rho > 0$. It is easy to observe from the description of wFA that as long as F is a fixed-size scheme, so is $\mathsf{wFA}^{\mathsf{F}}$. Moreover, for fixed-size F the scheme $\mathsf{wFA}^{\mathsf{F}}$ also achieves $\mathsf{w}_C(P) = 1$ with probability 1, as step (1) assigns seats of total weight $1 - \rho$, and step (2) assigns n_2 seats of weight ρ/n_2 each.

For brevity, let $\mathcal{W}$ and C denote the output distributions of $\mathsf{wFA}^{\mathsf{F}}(P, \mathcal{S}, n)$ and $\mathsf{F}(P, \mathcal{S}_3, n - n_1)$, respectively. Let $\tilde{A} \subset P$ be the set witnessing $\overline{\mathsf{E}}^{\mathcal{S}}_{n,\epsilon}[\mathsf{wFA}^{\mathsf{F}}]$, i.e., the one maximizing the probability

$$\Pr_{(C,\mathsf{w}_C)\leftarrow\mathcal{W}}\left[\mathsf{w}_C(\tilde{A}) \geq \mathcal{S}(\tilde{A}) + \epsilon\right] .$$

As in the description of wFA in Figure 6, for convenience we will assume (without loss of generality) that the parties in P are indexed in a non-increasing order with respect to their stake in $\mathcal{S}$, i.e., $i < j \Rightarrow \mathcal{S}(p_i) \geq \mathcal{S}(p_j)$.

Consider a partition of the party set P into two subsets P_1 and P_2, where $P_1 = \{p_i \in P \mid i \leq n_1\}$ and $P_2 = \{p_i \in P \mid i > n_1\}$. The crucial property motivating this partitioning is that while parties in P_1 are deterministically assigned a seat on the committee in step (1) of the scheme wFA in Fig. 6, for parties in P_2 their memberhip on the committee is decided later in step (2) and is deferred to the fallback scheme F. Denote $\tilde{A}_i = \tilde{A} \cap P_i$ for both $i \in \{1, 2\}$, hence $\tilde{A}_1 \cup \tilde{A}_2$ is a partition of $\tilde{A}$.

We first observe that for any $p_i \in P_1$, according to step (1) of wFA, p_i will be assigned exactly one committee seat c_i with $\mathsf{w}_C(i) = \mathcal{S}(p_i)$. Therefore, for any $(C, \mathsf{w}_C) \leftarrow \mathcal{W}$ with $C = (c_1, \ldots, c_{|C|})$, we have

$$\mathsf{w}_C(\tilde{A}_1) = \sum_{i\colon c_i \in \tilde{A}_1} \mathcal{S}(c_i) = \mathcal{S}(\tilde{A}_1) . \tag{17}$$

On the other hand, notice that for any $p \in P_2$, (15) gives us $\mathcal{S}_3(p) = \mathcal{S}(p)/\rho$ and hence

$$\mathcal{S}_3(\tilde{A}_2) = \mathcal{S}(\tilde{A}_2)/\rho . \tag{18}$$

Using these observations, we finally get

$$\begin{aligned}
\overline{\mathsf{E}}^{\mathcal{S}}_{n,\epsilon}\left[\mathsf{wFA}^{\mathsf{F}}\right] &\overset{(a)}{=} \Pr_{(C,\mathsf{w}_C)\leftarrow\mathcal{W}}\left[\mathsf{w}_C(\tilde{A}) \geq \mathcal{S}(\tilde{A}) + \epsilon\right] \\
&\overset{(b)}{=} \Pr_{(C,\mathsf{w}_C)\leftarrow\mathcal{W}}\left[\mathsf{w}_C(\tilde{A}_2) \geq \mathcal{S}(\tilde{A}_2) + \epsilon\right] \\
&\overset{(c)}{=} \Pr_{C'\leftarrow C}\left[\frac{\rho}{|C'|} \cdot \#_{\tilde{A}_2}(C') \geq \mathcal{S}_3(\tilde{A}_2)\rho + \epsilon\right] \\
&= \Pr_{C'\leftarrow C}\left[\#_{\tilde{A}_2}(C') \geq \left(\mathcal{S}_3(\tilde{A}_2) + \frac{\epsilon}{\rho}\right) \cdot |C'|\right] \\
&\leq \mathsf{E}^{\mathcal{S}_3}_{n-n_1,\epsilon/\rho}[\mathsf{F}] .
\end{aligned}$$

The above computation is justified as follows: (a) comes from the definition of $\tilde{A}$ and the fact that $\mathsf{w}_C(P) = 1$; while (b) uses (17) and the fact that $\tilde{A}$ can be partitioned as $\tilde{A}_1 \cup \tilde{A}_2$. The interesting step is (c): as we know that the committee-membership (and weight) of any party in $\tilde{A}_2$ is decided in step (2) of wFA via F, we can restrict our attention to the output of F, and hence the distribution C. Moreover, note that any party assigned a seat c_i in step (2) of wFA is assigned weight $\mathsf{w}_C(i) = \rho/n_2$, hence we have

$$\mathsf{w}_C(\tilde{A}_2) = \frac{\rho}{n_2} \cdot \#_{\tilde{A}_2}(C') = \frac{\rho}{|C'|} \cdot \#_{\tilde{A}_2}(C') ,$$

as the size of the committee C' (output by F) is n_2, since F is a fixed-size scheme.. Moreover, step (c) also employs (18). This concludes the proof of the lemma. □

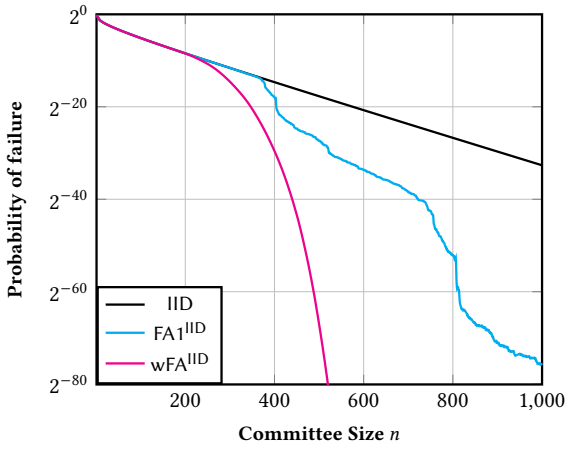

Figure 7: Cardano, $\epsilon = 0.1$.

It is now easy to establish Theorem 4.2 by directly applying Lemma 4.3 to $\mathsf{wFA}^{\mathsf{IID}}$, and subsequently invoking Lemmas 2.2 for IID.

5 EVALUATION

To provide evidence of the practical benefits achieved by our approach to committee-selection, in this section we evaluate our schemes on several real-world stake distributions of major proof-of-stake blockchains.

5.1 Stake distribution data

We have collected real-world stake distributions from three major proof-of-stake blockchains: Ethereum, Cardano, and Solana. These projects are respectively the second, seventh, and tenth largest cryptocurrencies by market capitalization at the time of writing,[1] and represent some of the largest and most decentralized proof-of-stake projects. Additionally, we also consider the stake distribution of Algorand, and a variant of the Cardano stake distribution with pools aggregated by owner, for reasons discussed in Section 5.3. Some basic statistics and inequality measures about the stake distributions we collected are summarized in Table 1.

To illustrate the effect of changing ϵ, in Figures 9 and 10 we provide two additional plots for Ethereum, analogous to Figure 1 but using $\epsilon = 0.05$ and $\epsilon = 0.2$, respectively.

We remark that due to the idiosyncrasies of the considered blockchains and their staking mechanisms, the data we use describe a slightly different population within each ecosystem, with the intention to always capture the "distribution of power" in the consensus mechanism. In other words, we select the stake distribution that is used by the consensus mechanism to attribute security-critical roles in the protocol, and this is typically the distribution that must be assumed to contain a certain fraction of honest stake for the protocol's security argument.

Note that while some of the considered blockchains contain committee selection as an integral part of their consensus protocols, and in these protocols the committee-selection part can be immediately

[1] Source: https://coinmarketcap.com.

Blockchain	Population size	Variation Coefficient	Gini coefficient	Theil index
Ethereum (ETH)	8674	31.59	0.95	5.2
Cardano (ADA)	3270	2.11	0.82	1.46
Cardano (ADA), aggregated	2028	7.22	0.92	2.62
Solana (SOL)	2443	3.6	0.76	1.64
Algorand (ALGO)	186	2.07	0.82	1.46

Table 1: Basic statistics about the considered real-world stake distributions.

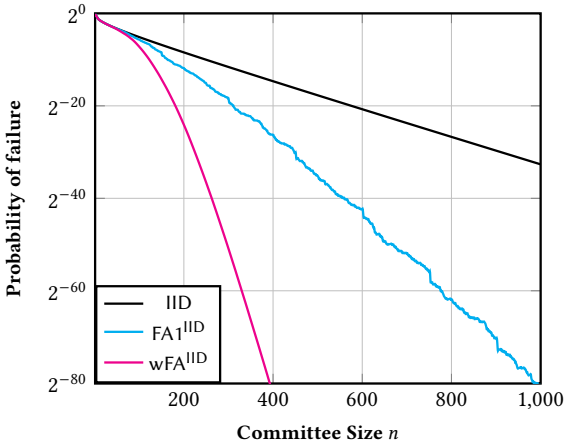

Figure 8: Solana, $\epsilon = 0.1$.

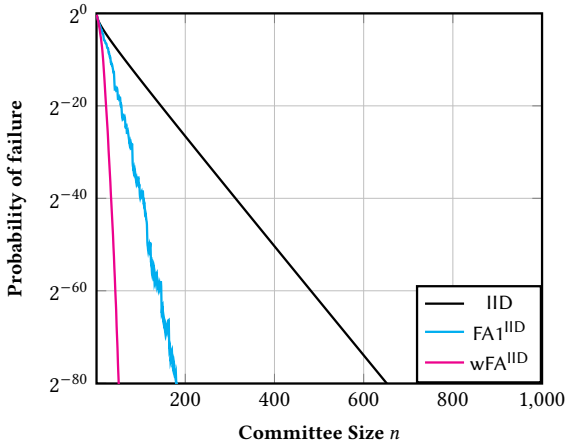

Figure 10: Ethereum, $\epsilon = 0.2$.

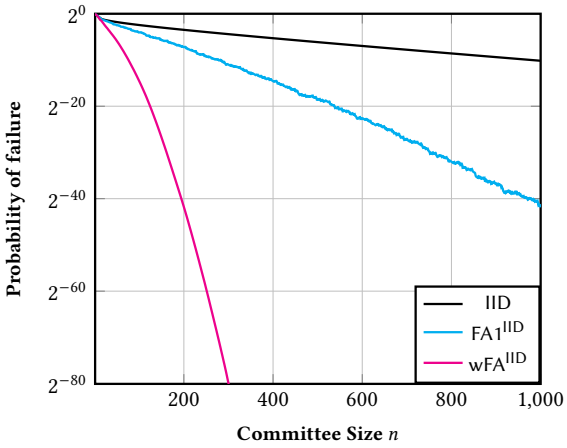

Figure 9: Ethereum, $\epsilon = 0.05$.

improved by our schemes, other considered blockchains do not employ a consensus of this type. Nonetheless, the reason why we consider these blockchains is that any layer-two committee-run protocol on top of any of these blockchains, if designed so as to derive its security from the same honest-majority assumption on stake that the underlying blockchain is making, would be sampling its committee from the respective distribution. This is the case irrespectively of whether the underlying layer-one consensus directly involves committee-sampling.

Concretely, for Ethereum the data describes the population of validators aggregated by *entities* running them, along with the normalized distribution of amounts they stake (each validator is staking exactly 32 ETH). For Cardano, we consider the population of the so-called *stake pools operators (SPOs)*, which are the parties running stake pools and participating in the consensus protocol by creating blocks. The stake distribution here is the normalized distribution of stake "delegated" to these pools by regular holders of ADA. For Solana, we consider the total population of block-producing *validators*, along with the normalized distribution of amounts they stake. Finally, for Algorand we consider the population of accounts that were online and participating in the consensus over a 7-day period, along with the normalized stake distribution.

To obtain the stake distribution data, we ran Selenium crawlers of web-based explorers for the considered chains and to query Dune analytics. The aggregation of Ethereum validators into entities is taken from Dune.[2] The Cardano data is obtained from their dbSync tool, and the aggregation of Cardano pools by owner is using cexplorer's group identification heuristics.[3] The data for Solana

[2]https://dune.com/queries/1933086/3188561
[3]https:/cexplorer.io

and Algorand come from a web-based explorer. [4] [5] All the data was collected in the first half of March, 2023.

5.2 Methodology

We upper-bound the error of our committee-selection schemes on each considered stake distribution $\mathcal{S}$ as follows:

- IID: We exactly evaluate the upper bound on $\mathsf{E}^{\mathcal{S}}_{n,\epsilon}[\mathsf{IID}]$ established in Lemma 2.3. To do this efficiently, observe that due to the properties of the Binomial distribution CDF, it suffices to consider at most $n+1$ values of θ to find the supremum in (6), rather than the whole interval $[0, 1]$. Note that this bound is independent of the distribution $\mathcal{S}$.
- LS: We evaluate the bound (7) obtained in Lemma 2.4. However, as we only evaluate LS in the context of Algorand which requires a below-1/3 corruption threshold (in the committee), we make the bounds more specific by upper-bounding the adversarial corruption α (in the underlying population) by $1/3 - \epsilon$. This bound is also independent of the distribution $\mathcal{S}$.
- $\mathsf{FA1}^{\mathsf{IID}}$: We exactly evaluate the bound (8) from Lemma 3.3. Namely, given the distribution $\mathcal{S}$, we determine $\mathcal{S}_\flat$, τ, and compute $\mathsf{E}^{(\mathcal{S}-\mathcal{S}_\flat)/\tau}_{\tau n,\epsilon/\tau}[\mathsf{IID}]$ as above.
- $\mathsf{FA1}^{\mathsf{LS}}$: We exactly evaluate the bound (11) from Lemma 3.4.
- $\mathsf{wFA}^{\mathsf{IID}}$: Similarly, here we also evaluate the bound (16) obtained in Lemma 4.3. Namely, given the distribution $\mathcal{S}$, we determine n_1, σ, and $\mathcal{S}_3$ as in Fig. 6, and compute $\mathsf{E}^{\mathcal{S}_3}_{n-n_1,\epsilon/\sigma}[\mathsf{IID}]$ as above.

Note that all provided estimates are upper bounds of the studied errors of interest. We choose to provide upper bounds for two reasons: we believe they are in most cases sufficiently tight for practice, and expect that any prudent parametrization of a deployed system (e.g., the choice of committee size) would be done based on upper bounds of the failure probability.

5.3 Results

Our results for Ethereum, Cardano and Solana are presented in Figures 1, 7, and 8, respectively. More concretely, each of these figures depicts an upper bound on the failure probability $\mathsf{E}^{\mathcal{S}}_{n,\epsilon}[\mathsf{F}]$ for $\mathsf{F} \in \{\mathsf{IID}, \mathsf{FA1}^{\mathsf{IID}}\}$ and $\overline{\mathsf{E}}^{\mathcal{S}}_{n,\epsilon}[\mathsf{wFA}^{\mathsf{IID}}]$ for values of $n \in \{1, \ldots, 1000\}$ and $\epsilon = 0.1$, with $\mathcal{S}$ being the stake distribution of the respective blockchain (Ethereum, Cardano, or Solana) as described in Section 5.1, and the bounds are obtained as outlined in Section 5.2. To maintain a scale that allows for comparison of all considered methods, we only display data points from the interval $[2^{-80}, 1]$.

Algorand and local sortition. Algorand natively uses committee selection as a critical part of its consensus algorithm, employing local (so-called *private*) sortition to provide adaptive security. Despite the smaller population size, we hence choose Algorand to showcase the local-sortition variant of our algorithm FA1. In Figure 11 we depict upper bounds (as detailed in Section 5.2) on the failure probability of LS and $\mathsf{FA1}^{\mathsf{LS}}$, again for values of $n \in \{1, \ldots, 1000\}$ and $\epsilon = 0.1$. As the failure probability of LS vanishes relatively slowly with increasing n, Algorand uses committees of sizes $n \in \{500, \ldots, 6000\}$

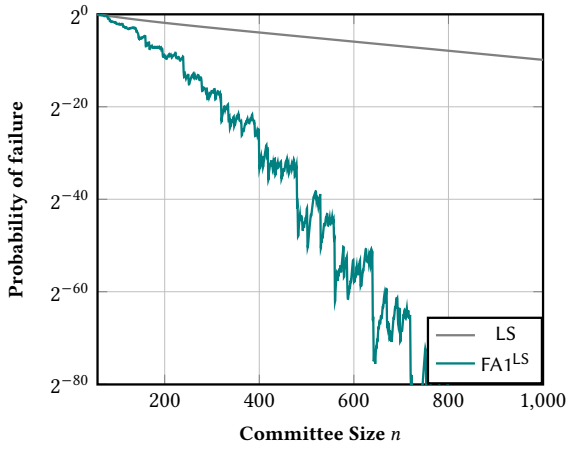

Figure 11: Algorand, $\epsilon = 0.1$.

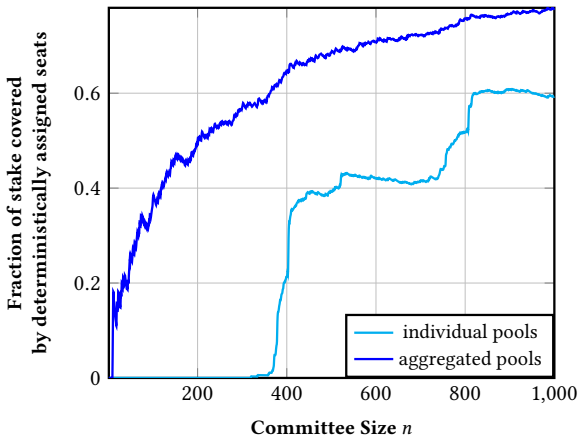

Figure 12: Fraction of stake assigned in step (1) of $\mathsf{FA1}^{\mathsf{IID}}$ applied to the Cardano stake distribution (individual pools vs. aggregated).

for various tasks within the protocol, depending on their criticality. [6] As Figure 11 illustrates, replacing LS by $\mathsf{FA1}^{\mathsf{LS}}$ would lead to comparable security levels with significantly smaller committees and hence significant savings in terms of network utilization for the consensus protocol (in Algorand each committee member has to disseminate a vote to all nodes).

Specifics of Cardano stake distribution; aggregation by owner. The error plot for $\mathsf{FA1}^{\mathsf{IID}}$ based on the Cardano stake distribution (in Fig. 7) looks distinctly different from the cases of Ethereum and Solana, as FA1 brings no advantage over the plain IID method up to almost $n = 400$. The reason for this is that Cardano has an incentive mechanism built into its reward-sharing scheme[7] that incentivizes stakeholders not to delegate to pools which already have roughly $1/400 = 0.25\%$ of stake delegated to them, as a measure to prevent

[4] https://solanabeach.io/validators
[5] https://algoexplorer.io/top-accounts

[6] See https://github.com/algorand/go-algorand/blob/master/config/consensus.go#L818.
[7] https://docs.cardano.org

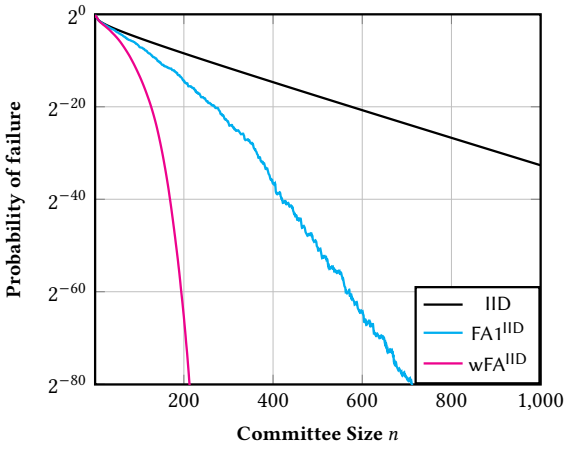

Figure 13: Cardano (aggregated by pool owner), $\epsilon = 0.1$.

centralization. Entities controlling more than this amount of stake typically resort to running several pools.

As a result, very few pools have more than 0.25% of stake delegated to them, and hence FA1 has very little to no effect for values of $n < 400$. We visualize this in Figure 12 which illustrates that FA1 does not assign almost any committee seats "deterministically" until $n \approx 400$.

To account for this point, we also provide an alternative plot for Cardano in Figure 13, where we use best-effort heuristics to aggregate pools that appear to have the same owner (based on metadata provided about the pool by its operator). This illustrates that if a reliable way to perform such aggregation existed within the ecosystem, this would significantly improve the performance of our schemes.

Acknowledgement. We would like to thank Stefan Contiu for collecting the datasets used for evaluating our schemes.

REFERENCES

[1] Alon Benhaim, Brett Hemenway Falk, and Gerry Tsoukalas. 2021. Scaling Blockchains: Can Elected Committees Help? *CoRR* abs/2110.08673 (2021). arXiv:2110.08673 https://arxiv.org/abs/2110.08673

[2] Ethan Buchman, Jae Kwon, and Zarko Milosevic. 2018. The latest gossip on BFT consensus. *arXiv preprint arXiv:1807.04938* (2018).

[3] Miguel Castro and Barbara Liskov. 2002. Practical Byzantine fault tolerance and proactive recovery. *ACM Transactions on Computer Systems (TOCS)* 20, 4 (2002), 398–461.

[4] Benjamin Y Chan and Elaine Shi. 2020. Streamlet: Textbook Streamlined Blockchains. Cryptology ePrint Archive, Report 2020/088. https://eprint.iacr.org/2020/088.

[5] Jing Chen and Silvio Micali. 2019. Algorand: A secure and efficient distributed ledger. *Theoretical Computer Science* 777 (2019), 155–183.

[6] Bernardo David, Peter Gazi, Aggelos Kiayias, and Alexander Russell. 2018. Ouroboros Praos: An Adaptively-Secure, Semi-synchronous Proof-of-Stake Blockchain. In *EUROCRYPT 2018, Part II (LNCS, Vol. 10821)*, Jesper Buus Nielsen and Vincent Rijmen (Eds.). Springer, Heidelberg, 66–98. https://doi.org/10.1007/978-3-319-78375-8_3

[7] Bernardo David, Bernardo Magri, Christian Matt, Jesper Buus Nielsen, and Daniel Tschudi. 2022. GearBox: Optimal-size Shard Committees by Leveraging the Safety-Liveness Dichotomy. In *ACM CCS 2022*, Heng Yin, Angelos Stavrou, Cas Cremers, and Elaine Shi (Eds.). ACM Press, 683–696. https://doi.org/10.1145/3548606.3559375

[8] Nicola Dimitri. 2022. The Economics of Consensus in Algorand. *FinTech* 1, 2 (2022), 164–179. https://doi.org/10.3390/fintech1020013

[9] Yossi Gilad, Rotem Hemo, Silvio Micali, Georgios Vlachos, and Nickolai Zeldovich. 2017. Algorand: Scaling Byzantine Agreements for Cryptocurrencies. In *Proceedings of the 26th Symposium on Operating Systems Principles* (Shanghai, China) *(SOSP '17)*. Association for Computing Machinery, New York, NY, USA, 51–68. https://doi.org/10.1145/3132747.3132757

[10] Wassily Hoeffding. 1994. Probability inequalities for sums of bounded random variables. In *The collected works of Wassily Hoeffding*. Springer, 409–426.

[11] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. 2017. Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol. In *CRYPTO 2017, Part I (LNCS, Vol. 10401)*, Jonathan Katz and Hovav Shacham (Eds.). Springer, Heidelberg, 357–388. https://doi.org/10.1007/978-3-319-63688-7_12

[12] S. Micali, M. Rabin, and S. Vadhan. 1999. Verifiable random functions. In *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*. 120–130. https://doi.org/10.1109/SFFCS.1999.814584

[13] Silvio Micali, Michael O. Rabin, and Salil P. Vadhan. 1999. Verifiable Random Functions. In *40th FOCS*. IEEE Computer Society Press, 120–130. https://doi.org/10.1109/SFFCS.1999.814584

[14] Andrew Miller, Yu Xia, Kyle Croman, Elaine Shi, and Dawn Song. 2016. The Honey Badger of BFT Protocols. In *ACM CCS 2016*, Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi (Eds.). ACM Press, 31–42. https://doi.org/10.1145/2976749.2978399

[15] Michael Mitzenmacher and Eli Upfal. 2005. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press.

[16] Maofan Yin, Dahlia Malkhi, Michael K. Reiter, Guy Golan-Gueta, and Ittai Abraham. 2019. HotStuff: BFT Consensus with Linearity and Responsiveness. In *38th ACM PODC*, Peter Robinson and Faith Ellen (Eds.). ACM, 347–356. https://doi.org/10.1145/3293611.3331591

A LARGE DEVIATION BOUNDS

Theorem A.1 (The Hoeffding bound). *Let $X_1, \ldots, X_n$ be a family of independent random variables taking values in $\{0, 1\}$. Let $S = \sum_i X_i - \mathbb{E}[X_i]$. Then for any $\lambda > 0$,*

$$\Pr[S \geq \lambda] \leq \exp(-2\lambda^2/n)\,.$$

Theorem A.2 (Poisson tail bound [15]). *Let P be a Poisson-distributed random variable with parameter λ. Then for any $x > \lambda$,*

$$\Pr[P \geq x] \leq \frac{(e\lambda)^x e^{-\lambda}}{x^x}$$

and for any $x < \lambda$

$$\Pr[P \leq x] \leq \frac{(e\lambda)^x e^{-\lambda}}{x^x}\,.$$

Theorem A.3 (Hoeffding's inequality, variance-dependent version; [10, (2.8)]). *Let $X_1, \ldots, X_n$ be independent random variables with zero mean and finite variance such that $X_i \leq 1$ (almost surely). Let $v = (1/n)\sum_i \mathbb{E}[X_i^2]$ and $S = \sum_i X_i$. Then for any $0 \leq \epsilon \leq 1$,*

$$\Pr[S \geq \epsilon n] \leq \left(\left(1 + \frac{\epsilon}{v}\right)^{-\frac{\epsilon+v}{1+v}} (1-\epsilon)^{-\frac{1-\epsilon}{1+v}}\right)^n\,.$$