COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 15 (3) (1995) 319–332

FAST PERFECT-INFORMATION LEADER-ELECTION PROTOCOLS WITH LINEAR IMMUNITY

JASON COOPER and NATHAN LINIAL*

Received February 26, 1993
Revised October 10, 1993

In this paper we develop a leader election protocol P with the following features:

1. The protocol runs in the *perfect information* model: Every step taken by a player is visible to all others.
2. It has *linear immunity*: If P is run by n players and a coalition of $c_1 n$ players deviates from the protocol, attempting to have one of them elected, their probability of success is $< 1 - c_2$, where $c_1, c_2 > 0$ are absolute constants.
3. It is *fast*: The running time of P is polylogarithmic in n, the number of players.

A previous protocol by Alon and Naor achieving linear immunity in the perfect information model has a linear time complexity. The main ingredient of our protocol is a *reduction* subprotocol. This is a way for n players to elect a subset of themselves which has the following property. Assume that up to εn of the players are bad and try to have as many of them elected to the subset. Then with high probability, the fraction of bad players among the elected ones will not exceed ε in a significant way. The existence of such a reduction protocol is first established by a probabilistic argument. Later an explicit construction is provided which is based on the spectral properties of Ramanujan graphs.

1. Introduction

A great deal of work has been dedicated to reliable computation in the presence of faults (see [9] for a recent survey). Among the most notable achievements of this line of research is work by Feldman and Micali [6] who present a randomized algorithm for Byzantine Agreement whose expected runtime is a constant. The main ingredient of their method is a randomized leader election protocol of constant expected run time, which is $(n-1)/3$-immune. Namely, even if any $t < n/3$ of the players may fail to follow the protocol, attempting to have one of them be elected, while everyone else does follow the protocol, and even if the failing players select the best coordinated strategy to achieve this goal, their probability of success is bounded away from 1. Like most theoretical work done on fault tolerance, Feldman and Micali postulate that data may be concealed. As usual, this is stated by either assuming completely secure communication channels, or by restricting the adversary's computational power. At the end of their paper they ask whether assumptions of this nature are necessary for the existence of such protocols.

Our understanding of fault-tolerant computation in a perfect-information environment, where all steps are visible to all players is not nearly as complete,

Mathematics Subject Classification (1991): 68 Q 20, 68 Q 22, 68 R 10, 05 C 90

* Work supported in part by grants from the Binational Science Foundation Israel–US and from the Israeli Academy of Sciences.

but see [4, 12, 8, 2, 1, 9]. In a perfect-information set-up, it is not even clear how to achieve linear immunity, let alone the protocol's time complexity. Perfect-information leader-election protocols of increasing immunity were found by Ben-Or Linial [4] and Saks in his baton-passing game [12] (see also [1]). Finally Alon and Naor [2] managed to achieve linear immunity. In the present paper we return to study the time complexity of such protocols, reducing the run time from Alon and Naor's $\Omega(n)$ (or even $\Omega(n^2)$ in the explicit version of their protocol). We are able to lower the time complexity to time poly-logarithmic in n, namely $\log^{17} n$. We speculate that a result analogous (and in a sense even stronger than) Feldman and Micali's may exist in this context, i.e. that there are perfect-information leader-election protocols with linear immunity which run in a constant number of steps. However, some new ideas will be required to achieve this goal. Note that [6] spend much effort to handle a dynamically forming bad coalition. In the present paper, in contrast, the bad coalition is fixed throughout, though it is selected in the worst possible way. To what extent such an adaptive adversary can be countered by perfect information protocols is still unclear.

The critical step in our method is a *reduction* protocol (e.g. [11]), where a set of n players jointly distributes $o(n)$ tokens among themselves. The number of tokens held by each player determines his voting power in future decisions. This process has the property that even if as many εn of the players may be cheaters, who fail to follow the protocol and attempt to collect as many tokens as possible, we can guarantee that with very high probability the fraction of tokens which go to cheaters does not significantly exceed ε. Given a fast reduction protocol it is fairly straightforward to achieve a fast leader-election protocol of linear immunity.

We first sketch a probabilistic proof of the existence of such 'reductions', and show how they can be used to create a poly-logarithmic-time leader election protocol immune against linear-sized coalitions. We then provide an explicit construction of a reduction protocol.

Rather than talk about tokens being distributed among the players, we sometime talk about a player getting a vote or being elected. When such terminology is employed, a player may get more than one vote, thus increasing his potential power in the future. Also, by an abuse of language we often talk about subsets of players, where indeed we have *multi-subsets* in mind, in which elements appear with multiplicities. Similarly, when we talk about committees that are being formed, it is the case that the same player may have more than one vote in the decisions made by such a committee.

2. Efficient reductions exist

In this section we show that a set of n players, some of whom may be cheaters, can be reduced to a smaller subset (indeed, sub-multiset) in such a way that the fraction of cheaters is 'unlikely' to grow 'significantly'.

Definition. An $(n, f, \varepsilon, \theta, \bar{\varepsilon})$-*reducer* is a protocol P performed by a set X of n players which, upon completion, distributes fn tokens among the players, and which has the following properties: Assume that the players in some unknown subset S of εn members in X are cheaters who may fail to follow P, and select their steps according to any coordinated strategy. Still, with probability at least $1-\theta$ no more than $(\varepsilon+\bar{\varepsilon})fn$ tokens go to members of S.

Theorem 2.1. *There is an absolute* $\varepsilon_0>0$, *so that for any* $\varepsilon<\varepsilon_0$ *and* $n>N(\varepsilon)$, *an* $\left(n, 1/\log^2 n, \varepsilon, e^{-\frac{n}{\log^{8.003} n}}, 1/\log^2 n\right)$*-reducer exists with running time* $O(\log^{16.01} n)$.

Remark. This is only an existence theorem. Specific constructions are presented later on.

Since this existence theorem does not yield results that are significantly stronger than our explicit construction, the proof will only be outlined, with an emphasis on ideas that appear in the construction as well. In the following sections we will need some results from [2], which we now outline:

Consider situations where many players run a protocol to make a collective choice. A large class of protocols can be described by a binary tree in the following way: label the internal nodes of the tree with names of players, and label the leaves with possible 'outcomes'. For the associated protocol, the player whose name labels the root flips a fair coin to pick a left or a right turn, and the process proceeds down the tree in this manner until an 'outcome' at a leaf is reached. We introduce the possibility of some subset of the players favoring a certain subset of the outcomes and making a concerted effort to direct the process to one of these favored outcomes. We call these players 'cheaters'. A cheater, when called upon to flip a coin, will pick a turn in the tree that will maximize the probability of a favored outcome being reached. The actions of the cheaters are coordinated according to an optimal strategy to bring about a desired result. Often we want to prove upper bounds on the probability of a favored outcome being reached. For a labeled tree T and a set of cheaters S we denote this probability by $P^T(S)$. Suppose that of n players εn are cheaters. A random labeling of the internal nodes of the tree is for all intents and purposes equivalent to labeling a node 'b' (for bad) with probability ε and 'g' (for guess what) with probability $1-\varepsilon$. Suppose also that a is the fraction of favored outcomes. A random labeling of the leaves is equivalent to labeling them 1 (for favored) with probability a and 0 with probability $1-a$. With every labeled tree T one associates the expectation φ_T of the outcome in T, averaging over all coin flips of non-cheaters. View φ as random variable on the probability space of all labeled trees, where all labelings are equally likely. We need bounds for the expectation and variance of this random variable. Letting d be the depth of the tree, [2] bound the expectation of φ by

$$E_T(\varphi_T) \leq a + \varepsilon \frac{\sqrt{a}}{\sqrt{2}(1-\sqrt{1/2+3\varepsilon/2})}$$

and its variance by

$$var(\varphi) \leq a\left(\frac{1}{2}+\frac{3\varepsilon}{2}\right)^d.$$

Thus, by Chebyschev's inequality, for any S and any positive γ, the fraction of labeled trees T for which

$$P^T(S) \geq a + \varepsilon \frac{\sqrt{a}}{\sqrt{2}(1-\sqrt{1/2+3\varepsilon/2})} + \gamma$$

is less than

$$\frac{a\left(\frac{1}{2}+\frac{3\varepsilon}{2}\right)^d}{\gamma^2}.$$

Assume now that the outcome is the choice of a leader, so leaves are labeled by names of players. The favored outcomes correspond to the election of a cheater, so $a{=}\varepsilon$. For any ε, γ and d large enough $(d{=}\Omega(n))$, the above probability is less than $1/\binom{n}{\varepsilon n}$ so summing over all coalitions of εn cheaters we get: if

$$\varepsilon + \frac{\varepsilon^{3/2}}{\sqrt{2}(1-\sqrt{1/2+3\varepsilon/2})} < 1$$

(e.g. $\varepsilon < 1/9$) then there is a leader-election protocol of run time $O(n)$ (depth of the tree) such that the probability of a cheater being elected is less than

$$\varepsilon + \frac{\varepsilon^{3/2}}{\sqrt{2}(1-\sqrt{1/2+3\varepsilon/2})} < \varepsilon + 4\varepsilon^{3/2}.$$

We are now ready to outline our existence proof: Consider a set of n players, and a fixed set of cheaters S with $|S| \leq \varepsilon n$. We are interested in the following class of protocols: Committees are constructed from the n players. The protocol consists of each committee selecting a subcommittee (or rather - distributing tokens). The 'reduction' is the multi-union of all selected sub-committees (all the players holding tokens, adding multiplicities). Distributing tokens in a committee is an instance of a choice problem, so various protocols can be given in terms of binary trees, as described above. Internal nodes will be labeled by names of players in the committee, and leaves will be labeled by possible distributions of tokens. We define:

Definition. An AN-tree on k players is a full binary tree of depth k^2, where the internal nodes are labeled by names of players, and the leaves are labeled by (multi) subsets of the players of size $\sqrt{k}$. The associated 'tree game' picks a subset in the usual way.

We can now describe the random construction. Note: To keep the exponents integral, and to allow the result to extend to our explicit construction, we will prove the following:

Theorem 2.2. *There is an absolute $\varepsilon_0 > 0$, so that for any $\varepsilon < \varepsilon_0$ and $n > N(\varepsilon)$, an $\left(n, 1/\log^2 n, \varepsilon, e^{-\frac{n}{\log^{11} n}}, 1/\log^2 n\right)$-reducer exists with running time $\Theta(\log^{52} n)$ (i.e. $n/\log^2 n$ tokens are distributed by n players among themselves, any εn of the players may be cheaters, but still, with probability $\geq 1 - e^{-\frac{n}{\log^{11} n}}$ no more than $(\varepsilon + 1/\log^2 n)(n/log^2 n)$ tokens will go to cheaters).*

The random construction is as follows: Select $n/\log^{15} n$ committees from the n players, each of size $\log^{26} n$. These committees are selected independently, uniformly, at random with repetitions. (In particular, a player may appear in many committees, and likewise, may appear in a committee more than once, in which case his different occurrences are regarded as different players, as previously remarked.)

For each committee select a random AN-tree by which it will select a sub-committee of $\log^{13} n$ players. Internal labels are picked independently, uniformly, at random with repetitions. Members in the sub-committees at the leaves are selected likewise. Call a committee *bad* if the fraction of cheaters in it exceeds $\varepsilon + 1/\log^4 n$. By a well known large-deviation theorem ([3], for one), the probability of a particular committee being *bad* is no greater than

$$(1) \qquad e^{-2(1/\log^4 n)^2 \log^{26} n} = e^{-2\log^{18} n}.$$

Similarly, call a sub-committee *bad* if the fraction of cheaters in it is greater than $\varepsilon + 1/\log^3 n$. Let a be the probability of a given sub-committee being *bad*, conditioned on its being a sub-committee of a *good* committee. By the same theorem,

$$(2) \qquad a < e^{-2(1/\log^3 n - 1/\log^4 n)^2 \log^{26/2} n} \leq e^{-\log^7 n}$$

for sufficiently large n. Note that for a *good* committee, our random AN-tree is exactly of the type discussed at the beginning of this section, where the internal nodes are bad with probability at most $\varepsilon + 1/\log^4 n$ and the leaves are bad with probability at most a. Call the AN-tree *bad* if the probability (over the coin flips) of a *bad* sub-committee being selected is greater than

$$(3) \qquad a + \left(\varepsilon + \frac{1}{\log^4 n}\right) \frac{\sqrt{a/2}}{\left(1 - \sqrt{1/2 + 3\left(\varepsilon + \frac{1}{\log^4 n}\right)/2}\right)} + \gamma$$

for some as yet undefined γ. As mentioned above, for any γ, the probability (over the random choice of the AN-tree) of a tree being *bad* does not exceed

$$(4) \qquad \frac{a(1/2 + 3(\varepsilon + 1/\log^4 n)/2)^d}{\gamma^2}$$

where $d = \log^{52} n$ is the depth of the tree. Note that most committees thus constructed will be *good*, and for most *good* committees a *good* tree will be chosen. We calculate a bound on the probability of things going seriously wrong.

Of our $n/\log^{15} n$ committees, let α be the fraction of them in which either the committee itself or the AN-tree assigned to it is *bad*. Let β be the fraction of them in which the committee and the tree are *good*, but due to bad luck, a *bad* sub-committee was chosen anyway. Assume the worst: all *bad* sub-committees consist only of cheaters. Assume also that *bad* committees and committees using *bad* trees always choose *bad* sub-committees. The fraction of cheaters in the reduced set will be at most

$$(\alpha + \beta) + [1 - (\alpha + \beta)](\varepsilon + 1/\log^3 n) = (\alpha + \beta)[1 - (\varepsilon + 1/\log^3 n)] + \varepsilon + 1/\log^3 n$$

Our reduction fails if the above exceeds $\varepsilon + 1/\log^2 n$, that is, if

$$(\alpha + \beta) > \frac{(1/\log^2 n - 1/\log^3 n)}{[1 - (\varepsilon + 1/\log^3 n)]}.$$

Clearly, for this to happen either α or β would have to be $\Omega(1/\log^2 n)$. To calculate bounds on the probability of this happening, we set γ to be $e^{-\log^{51} n}$.

The probability of a committee being *bad* is, by (1), at most $e^{-2\log^{18} n}$, and the probability of a tree being *bad*, given a *good* committee, is, by (4), at most

$$\frac{a(1/2+3(\varepsilon+1/\log^4 n)/2)^d}{e^{-2\log^{51} n}}$$

which is $O(e^{-\log^{51} n})$ (for a sufficiently small, but predetermined, ε). Note that the probability of a *bad* committee is overwhelmingly greater than the probability of a *bad* tree. By the Hoeffding inequality, the probability of $\alpha=\Omega(1/\log^2 n)$ is at most

$$((c\,e^{-2\log^{18} n}(\log^2 n))^{1/c\log^2 n})^{\frac{n}{\log^{15} n}}$$

(for some constant c), which is $e^{-\Theta(n\log n)}$, and in any event is much smaller than

$$\frac{1}{\binom{n}{\varepsilon n}}$$

for sufficiently large n. Summing over all

$$\binom{n}{\varepsilon n}$$

choices of εn cheaters we conclude that there exists a selection of committees and trees for which α is significantly less than $1/\log^2 n$, for any set of εn cheaters.

We now calculate a bound on the probability of β being $\Omega(1/\log^2 n)$. The probability of a *good* committee picking a *bad* sub-committee using a *good* tree is, by definition (3), no more than

$$a+\frac{\left(\varepsilon+\frac{1}{\log^4 n}\right)\sqrt{a}}{\sqrt{2}\left(1-\sqrt{\frac{1}{2}+\frac{3\left(\varepsilon+\frac{1}{\log^4 n}\right)}{2}}\right)}+e^{-\log^{51} n}\leq$$

$$e^{-\log^7 n}+\frac{\left(\varepsilon+\frac{1}{\log^4 n}\right)e^{-\frac{1}{2}\log^7 n}}{\sqrt{2}\left(1-\sqrt{\frac{1}{2}+\frac{3\left(\varepsilon+\frac{1}{\log^4 n}\right)}{2}}\right)}+e^{-\log^{51} n}$$

which is $O\left(e^{-\frac{1}{2}\log^7 n}\right)$. Again by the Hoeffding inequality, the probability that $\beta=\Omega(1/\log^2 n)$ is at most

$$((ce^{-\log^6 n}(\log^2 n))^{1/c\log^2 n})^{\frac{n}{\log^{15} n}}$$

(for some constant c), which is no greater than $e^{-\frac{n}{\log^{11} n}}$ for sufficiently large n.

All that remains is to evaluate the running time. The described protocol is of running time $\Theta(\log^{52} n)$, since each committee runs through a tree of depth $\log^{52} n$, in parallel.

Note. The better result stated at the beginning of this section can be achieved by modifying the construction as follows: Select $n/\log^{10.005} n$ committees from the n players, each of size $\log^{16.01} n$. For each committee select a random AN-tree of depth $O(\log^{16.01} n)$. Call a committee *bad* if the fraction of cheaters in it exceeds $\varepsilon + 1/\log^{2.002} n$. Call a sub-committee *bad* if the fraction of cheaters in it exceeds $\varepsilon + 1/\log^{2.001} n$. The given proof works for these smaller exponents.

3. The existence of leader-election protocols

Theorem 3.1. $\left(n, 1/\log^2 n, \varepsilon, e^{-\frac{n}{\log^{11} n}}, 1/\log^2 n\right)$*-reducers with running time T may be used to get a leader-election protocol immune against linear sized coalitions, with running time $O(T \log n)$.*

This, together with the results of the previous (and the following) section, yields the following result:

Theorem 3.2. *There are $\mu > 0$, $c < 1$, and there is an explicitly constructed leader election protocol with running time poly-logarithmic in n, such that for any $\varepsilon \leq \mu$, when the protocol is run on n players, at most εn of whom are cheaters, the probability of a cheater being elected is less than c.*

Remark. Explicitness is postponed to the following section.

Proof. Given the existence of reductions from the previous section, constructing a protocol is easy: reduce the n players down to $n/\log^2 n$, and keep on reducing until you reach some threshold n_0 (independent of n). On these n_0 players play the baton-passing game [1], and let the leader be the player thus chosen. All we need to do is select the threshold, and show that the running time is indeed $O(T \log n)$, and that the probability of electing a cheater is bounded away from 1.

First we evaluate the running time, for any threshold n_0. Let us first count the number of reductions needed to come down from n to $\sqrt{n}$ players. Letting $k = \log n$, we are coming down from 2^k players to $2^{k/2}$. A reduction protocol performed by 2^l players brings their number down to $2^{l-2\log l}$. For $\frac{1}{2}k \leq l \leq k$, a reduction step by 2^l players results in a set of $2^{l-2\log l} \leq 2^{l-2\log k/2} = 2^{l-2\log k+2}$ players. So a set of 2^k players reduces in size to $2^{k/2}$ in no more than $\frac{k}{4(\log k - 1)} = \frac{\log n}{4(\log\log n - 1)}$ steps, and certainly no more than $\log n$. Letting $n = n_0^{2^t}$, the number of reductions needed till there are no more than n_0 players is at most

$$(\log n_0)(2 + 4 + 8 + \ldots + 2^t) \leq 2(\log n_0)(\log n).$$

Since each reduction takes time $O(T)$, the time needed to reduce n players to n_0 is $O(T \log n)$.

Let us now see what happens to the population of cheaters as we keep reducing the number of players. Call a reduction on m players successful if the fraction of cheaters increases by no more than $1/\log^2 m$. Starting with n players, εn of whom are cheaters, and assuming that all the reductions on the way down to n_0 players are successful, how high can the fraction of cheaters be at the end? Following the analysis in the previous paragraph, we first estimate the increase in the fraction of cheaters as the total number of players reduces from n to $\sqrt{n}$. By assumption, in this range of parameters, the fraction of cheaters will increase by no more than $1/\log^2\sqrt{n}=4/\log^2 n$ per reduction step, the greatest possible increase being at the last round. Recall that at most $\frac{\log n}{4(\log\log n-1)}$ reductions are needed to decrease from n to $\sqrt{n}$ players, the fraction of cheaters will grow by no more than $\frac{1}{\log n(\log\log n-1)}$. Again let $n=n_0^{2^t}$, the total increase in the fraction of cheaters will be at most

$$\frac{1}{2\log n_0\ \log\log n_0}+\frac{1}{4\log n_0(\log\log n_0+1)}+\ldots+\frac{1}{2^t\log n_0(\log\log n_0+t-1)}$$
$$\leq\frac{1}{\log n_0\log\log n_0}(1/2+1/4+\cdots)=\frac{1}{\log n_0\ \log\log n_0}.$$

This consideration was assuming all the reductions to be successful. To bound the probability of this not being the case, we must sum the probabilities of unsuccessful reductions. Recalling that the probability of an unsuccessful reduction on z players, for sufficiently large z, is no greater than $e^{-\frac{z}{\log^{11} z}}$, the probability of at least one unsuccessful reduction coming down from z to $\sqrt{z}$ is at most $\frac{\log z}{4(\log\log z-1)}e^{-\frac{\sqrt{z}}{\log^{11}\sqrt{z}}}$, which is less than $e^{-z^{1/3}}$ for sufficiently large z, and summing over all z of the form $n_0^{2^i}$, $i\geq 1$ we get

$$e^{-(n_0^{1/3})}+e^{-(n_0^{1/3})^2}+e^{-(n_0^{1/3})^4}+e^{-(n_0^{1/3})^8}+\ldots$$

which clearly gives an arbitrarily low probability of an unsuccessful reduction, depending on the choice of n_0.

What remains now is to pick $n_0=n_0(\varepsilon)$ in such a way that there will be absolute constants μ and $c<1$ such that for any $\varepsilon\leq\mu$ our protocol will elect a cheater with probability at most c. The last stage of our protocol is baton-passing on n_0 players where the fraction of cheaters is 'probably' no greater than $\varepsilon+\frac{1}{\log n_0\log\log n_0}$. Assuming that is the case, for $\log n_0\log\log n_0\geq 1/\varepsilon$, the fraction of cheaters will be no greater than 2ε. By the analysis in [1], the probability of electing a cheater will be bounded by 20ε, as long as $2\varepsilon\leq\frac{1}{3\log n_0}$, that is to say,

$$\log n_0\ \log\log n_0\leq\frac{1}{6\varepsilon}\log(1/6\varepsilon).$$

The two conditions on n_0 can hold simultaneously as long as $\varepsilon\leq\frac{1}{6e^6}$. So we take n_0 to be $e^{1/6\varepsilon}$. We bound the probability of selecting a cheater by the sum of two probabilities: that of the final stage electing a cheater, assuming we did indeed end up with no more than $2\varepsilon n_0$ cheaters, and that of ending up with more than $2\varepsilon n_0$ cheaters at the final stage. The first is bounded by 20ε, and the second, if n_0 is large enough for the results of Section 2 to hold, by (5), which can be made arbitrarily small letting n_0 be sufficiently large (that is — letting ε be sufficiently small). So bounding ε by a sufficiently small μ bounds the probability of choosing a cheater away from 1, and the proof is complete. ∎

4. Constructions

The overall structure of our protocol is this: We provide an explicit construction for reducers. Namely, we show an explicit family of $n/\log^{12} n$ subsets of $[n]$ ("committees") of size $\log^{18} n$ each, with the property that for every subset S of εn cheaters, in no more than $O\left(\frac{\varepsilon n}{\log^{14} n}\right)$ committees does the fraction of cheaters exceed $\varepsilon + 1/\log^2 n$. Now, using the method of [2] we let each committee select a sub-committee of $\log^9 n$ members in poly-logarithmic time. A *good* committee is very unlikely to select a sub-committee in which the fraction of cheaters exceeds $\varepsilon + 1/\log^2 n$. The process is then repeated with the (multi-) union of selected sub-committees. As described in Section 3, a logarithmic number of iterations reduces the set of n players to a subset of some constant size (dependent on ε), and with very high probability does not increase the fraction of cheaters by more than $2/\log^2 n$. The details of the analysis follow.

First we describe a committee selection, which is based on set-systems that are good dispersers (for a general reference on dispersers, see [5]). For a family $\mathcal{F}$ of subsets of $[n]$, a parameter $\delta > 0$, and a subset $S \subseteq [n]$, let $\mathcal{F}(\mathrm{S},\delta)$ be the family of all $F \in \mathcal{F}$ for which

$$\frac{|F \cap S|}{|F|} > \frac{|S|}{n} + \delta.$$

Proposition 4.1. *There is a family $\mathcal{F}$ of $n/\log^{12} n$ subsets of $[n]$, each of size $\log^{18} n$, such that for any subset S of $[n]$ of size εn, and any δ,*

$$\frac{|\mathcal{F}(S,\delta)|}{|\mathcal{F}|} = O\left(\frac{\varepsilon}{\delta^2 \log^6 n}\right).$$

Proof. We represent the family $\mathcal{F}$ by a matrix A whose rows correspond to players and columns to committees, with $a_{ij} = 1$ if player i is in committee j, and 0 otherwise. Here is the construction: Consider a Ramanujan graph $G = (V,E)$ as in [10], and remove edges from it, so it becomes regular of degree $d = (q^2 + q + 1)$ for q a prime power. By standard results on the density of primes, and by the variational formula for eigenvalues, there is a choice of an initial Ramanujan graph so that the second eigenvalue of the graph thus obtained is still $O(d^{0.75})$. Also, the choice can be made to guarantee that $girth(G)$ be arbitrarily large (≥ 5 will suffice). Note: Noga Alon (private communication) has indicated to us that, in fact, one may obtain a second eigenvalue of $O(\sqrt{d})$, but this will not affect our arguments.

Associate with each vertex $v \in V$ a projective plane P_v of order q. Each edge incident with vertex v is associated with one of the (q^2+q+1) lines in the plane. Rows in our matrix A correspond to pairs (v,p) where v is a vertex of G and p is a point in the plane P_v. The columns of A correspond to vertices of G. We can now describe our matrix A: $A_{(v,p),w}=1$ iff $[v,w]\in E(G)$ and the line in the plane P_v corresponding to that edge goes through point p. All other entries in A are 0. Since each point in the plane has $(q+1)$ lines through it, the sum of each row in A is exactly $(q+1)$, i.e. $A\mathbf{1}=(q+1)\mathbf{1}$ ($\mathbf{1}$ being the all 1's vector of the appropriate dimension). Also, the sum of each column in A is exactly $(q+1)(q^2+q+1)$, i.e. $\mathbf{1}A=(q+1)(q^2+q+1)\mathbf{1}$. This holds, since each vertex v has (q^2+q+1) neighbors in the graph, and for each neighbor w of v there are $(q+1)$ points in P_v on the line associated with the edge $[v,w]$. Other interesting properties of A are revealed by considering the matrix A^TA, in which the (v,w) entry is the dot product of columns v and w in A. For $v=w$ we get $(q+1)(q^2+q+1)$ (the number of 1's in a column). For $v\neq w$ the (z,p) entry in both corresponding columns equals 1 iff: (1) z is a neighbor of both v and w, and (2) the edges $[v,z]$ and $[w,z]$ both correspond to lines through p. Since $girth(G)\geq 5$, (1) holds exactly when the distance between v and w in G is 2, and in that case there is exactly one such z. Since every two lines in a projective plane have exactly one point in common, there is exactly one suitable p. Summing up, we see that

$$(A^TA)_{v,w}=\begin{cases}(q+1)(q^2+q+1) & \text{if } v=w\\ 1 & \text{if } d(v,w)=2\\ 0 & \text{otherwise.}\end{cases}$$

Let M be the adjacency matrix of graph G ($M_{v,w}=1$ if $[v,w]$ is an edge, and 0 otherwise). Clearly M^2 is as follows:

$$(M^2)_{v,w}=\begin{cases}d=q^2+q+1 & \text{if } v=w\\ 1 & \text{if } d(v,w)=2\\ 0 & \text{otherwise,}\end{cases}$$

so our statement regarding A^TA can be rephrased as

$$A^TA=M^2-(q^2+q+1)I+(q+1)(q^2+q+1)I=M^2+q(q^2+q+1)I.$$

As we mentioned, M's second eigenvalue is $O(q^{1.5})$, where $d=(q^2+q+1)$ is the first eigenvalue. The first and second eigenvalues of M^2 are therefore d^2 and $O(d^{1.5})$ respectively, and consequently, the first and second eigenvalues of A^TA are $\lambda_1=(q^2+q+1)^2+q(q^2+q+1)=(q^2+q+1)(q+1)^2$ and $\lambda_2=O(d^{1.5})$ respectively.

Our matrix A describes a collection of $|V|$ committees from $n=|V|(q^2+q+1)$ players, each of size $(q+1)(q^2+q+1)$. Consider a coalition of εn cheaters, whose characteristic vector u is a row vector of length n, ($u_i=1$ if player i is a cheater, and 0 otherwise). If $v=uA$, then for any i, v_i is the number of cheaters in the i-th committee. Ideally all committees would have fraction ε of cheaters, and all the entries in v would be equal. To gauge the deviation from this ideal situation we estimate v in both ℓ_1 and ℓ_2 norm. The average fraction of cheaters in a committee is ε, i.e.

$$\sum_{i=1}^{|V|} v_i = \langle v, \mathbf{1}\rangle = uA\mathbf{1} = (q+1)\langle u, \mathbf{1}\rangle = \varepsilon n(q+1) = \varepsilon |V|(q+1)(q^2+q+1).$$

We now calculate the ℓ_2 norm. Note that

$$\sum_{i=1}^{|V|} v_i^2 = vv^T = uAA^Tu^T.$$

Recall that the non zero eigenvalues of AA^T and A^TA are exactly the same. Let $\{w_i\}$ be an orthonormal base of eigenvectors of AA^T corresponding to eigenvalues $\{\lambda_i\}$, and let $u = \sum c_i w_i$ be the expansion of u in this base. Since u is a 0–1 vector, and the w_i are orthonormal, $\varepsilon n = \sum u_i = \sum u_i^2 = uu^T = \left(\sum c_i w_i\right)\left(\sum c_i w_i^T\right) = \sum c_i^2$. Furthermore, $c_1 = uw_1^T$, and $w_1 = (1/\sqrt{n})\mathbf{1}$, whence $c_1 = \varepsilon\sqrt{n}$. We are now ready to calculate $\sum v_i^2$:

$$\begin{aligned}
\sum v_i^2 &= vv^T = uAA^Tu^T = \left(\sum c_i w_i\right)(AA^T)\left(\sum c_j w_j^T\right) \\
&= \left(\sum c_i w_i\right)\left(\sum c_j \lambda_j w_j^T\right) = \sum_i c_i^2 \lambda_i = c_1^2\lambda_1 + \sum_{i>1} c_i^2 \lambda_i \\
&\leq \varepsilon^2 n\lambda_1 + \lambda_2 \sum c_i^2 = \varepsilon^2 n(q^2+q+1)(q+1)^2 + O(d^{1.5})\varepsilon n \\
&= \varepsilon^2 |V|(q^2+q+1)^2(q+1)^2 + O(\varepsilon |V| d^{2.5}).
\end{aligned}$$

To state our result regarding the committees, fix a set of εn cheaters with characteristic vector u. Consider the set of committees as a probability space with uniform distribution, and let X be the random variable that associates with every committee the fraction of cheaters in it. Clearly, $(q^2+q+1)(q+1)X$ is the random variable that associates with every committee the *number* of cheaters in it, so $E(X^2) = \frac{\|v\|_2^2}{|V|(q+1)^2(q^2+q+1)^2}$. As we have shown, X has expectation ε, so we can now estimate the variance:

$$\begin{aligned}
\sigma^2(X) &= E(X^2) - (E(X))^2 \\
&= \frac{\varepsilon^2|V|(q^2+q+1)^2(q+1)^2 + O(\varepsilon|V|d^{2.5})}{|V|(q+1)^2(q^2+q+1)^2} - \varepsilon^2 = O(\varepsilon d^{-1/2}).
\end{aligned}$$

We can now use Chebyschev's inequality to estimate the number of committees with large representation of cheaters:

$$\Pr(X > \varepsilon + \delta) < \frac{\sigma^2(X)}{\delta^2} = O\left(\frac{\varepsilon}{\delta^2\sqrt{d}}\right).$$

That is to say that the fraction of *bad* committees is $O\left(\frac{\varepsilon}{\delta^2\sqrt{d}}\right)$ where we call a committee *bad* if the fraction of cheaters in it exceeds $\varepsilon + \delta$, for some arbitrary δ.

We now show how to use the above described family of subsets to construct a reducer. What we need to show is how committees of size k select sub-committees of size $\sqrt{k}$ in time $O(k^2)$. To describe this stage we first recall some facts and results from [2], [12] and [1].

(i) Given k players, there is an explicit construction for selecting k committees of size $\sqrt{k}$ each, in such a way that for any set of εk cheaters, the number of committees in which the fraction of cheaters exceeds $(\varepsilon+\alpha)$ is at most $\frac{\varepsilon}{\alpha^2}\sqrt{k}$ for any threshold α.

(ii) There is an explicit leader election protocol on k players that runs in time $O(k^2)$, such that if the fraction of cheaters ε is sufficiently small, the probability of a cheater being elected is at most $c\varepsilon$ for some constant c.

(iii) In the baton-passing leader election game the first player receives a baton, and each player upon receiving it, passes it on to some player who has not yet held it. The leader is the last player to receive the baton.

In the faulty- baton-passing game on n players, εn of whom are cheaters, each good player (non cheater) has some positive probability θ of turning bad upon receiving the baton. The probability of a cheater being elected in this game is at most $8\frac{(\varepsilon n\log(\varepsilon n+1))^{\frac{1}{1-4\theta}}}{(n-\varepsilon n)^{1-\theta}}$.

Now back to the construction. Given k players, determine k sub-committees of size $\sqrt{k}$ each by (i). These sub-committees play the (faulty) baton-passing game as follows: The first sub-committee selects a leader using (ii) in time $O(k)$. The leader passes the baton on to an as yet unselected sub-committee, and so on until the baton reaches the last sub-committee. This last sub-committee is the selected one. The cheaters' strategy will, by the now famous moment's reflection argument, be for the cheaters in a sub-committee to try and get one of theirs elected (recursively), and if they succeed, he should pick a sub-committee containing the least number of cheaters. Now let μ be the fraction of cheaters in a committee, and call a sub-committee *bad* if the fraction of cheaters in it is greater than $\mu+\alpha$ for some threshold α. By (i) the number of *bad* committees will be at most

$$\frac{\mu}{\alpha^2}\sqrt{k}.$$

If we pessimistically (as usual) assume that *bad* sub-committees always elect a cheater, then what we are doing is playing 'faulty baton-passing' with sub-committees as players, where a good player turning bad corresponds to a *good* sub-committee electing a cheater. By (ii) there is some constant c such that the probability of this event is less than $c(\mu+\alpha)$, provided the sub-committees are large enough and $(\mu+\alpha)$ is small enough. We may assume this to be the case, and therefore we will assume that the probability of a good 'player' turning bad is less than θ for some arbitrarily small θ. By (iii), the probability of a *bad* sub-committee being chosen is bounded by

$$8\frac{\left(\frac{\mu}{\alpha^2}\sqrt{k}\log\left(\frac{\mu}{\alpha^2}\sqrt{k}+1\right)\right)^{\frac{1}{1-4\theta}}}{\left(n-\frac{\mu}{\alpha^2}\sqrt{k}\right)^{1-\theta}}.$$

We can now describe an explicit $(n, 1/\log^3 n, \varepsilon, 2^{-\frac{n}{\log^{14} n}}, 5/\log^2 n)$-reducer. Given n players of whom εn are cheaters, let d be the number closest (from below) to $\log^{12} n$ such that $d = (q^2 + q + 1)$ for q a prime power. d will be the degree of an expander graph G on $|V| = n/d$ vertices. This graph describes a selection of $|V| = \frac{n}{\log^{12} n}$ committees, each of size $d^{1.5} = \log^{18} n$. Each committee will select a sub-committee of size $\log^9 n$ as described above, and the union of these sub-committees will consist of $\frac{n}{\log^3 n}$ players. Call a committee *bad* if the fraction of cheaters in it exceeds $\varepsilon + 1/\log^2 n$. By Proposition 4.1, there are no more than

$$\frac{n}{\log^{12}} O\left(\frac{\varepsilon \log^4 n}{\log^6 n}\right) = O\left(\frac{\varepsilon n}{\log^{14} n}\right)$$

bad committees. Now consider a *good* committee, that is - one where the fraction of cheaters is less than $\varepsilon + 1/\log^2 n$. From it we generate $\log^{18} n$ sub-committees of size $\log^9 n$ each. Call a sub-committee *bad* if the fraction of cheaters in it is greater than $\varepsilon + 2/\log^2 n$. By (i) at most

$$\frac{\varepsilon + 1/\log^2 n}{(1/\log^2 n)^2} \sqrt{\log^{18} n} = (\varepsilon + 1/\log^2 n) \log^{13} n$$

sub-committees are *bad*. Let us estimate the probability of our *good* committee selecting a *bad* sub-committee. By (6) the probability is at most

$$\frac{8((\varepsilon + 1/\log^2 n)\log^{13} n \log((\varepsilon + 1/\log^2 n)\log^{13} n + 1))^{\frac{1}{1-4\theta}}}{(\log^{18} n - (\varepsilon + 1/\log^2 n)\log^{13} n)^{1-\theta}}$$

$$\leq c \frac{\log^{\frac{13(5\theta - 4\theta^2)}{1-4\theta}} n (\log\log n)^{1/(1-4\theta)}}{(\log^5 n)^{1-\theta}} \leq \frac{1}{\log^4 n}$$

for sufficiently large n. We now count the number of cheaters in the reduced set (which is of size $\frac{n}{\log^3 n}$). The *bad* committees can contribute at most

$$O\left(\frac{\varepsilon n}{\log^{14} n}\right) \log^9 n = \frac{n}{\log^3 n} O\left(\frac{\varepsilon}{\log^2 n}\right).$$

Suppose K *good* committees selected *bad* sub-committees. They will contribute at most $K \log^9 n$ cheaters. All the other committees together will contribute no more than

$$\frac{n}{\log^3 n}(\varepsilon + 2/\log^2 n).$$

The total number of cheaters selected is no more than

$$\frac{n}{\log^3 n}\left(\varepsilon + \frac{3}{\log^2 n} + K\frac{\log^{12} n}{n}\right).$$

This is no more than $\frac{n}{\log^3 n}(\varepsilon + 5/\log^2 n)$ as long as $K \leq \frac{n}{\log^{14} n}$.

By the Hoeffding inequality, the probability of this not being the case is

$$\Pr\left(K \geq \frac{n}{\log^{14} n}\right) = \Pr\left(K \geq \frac{n}{\log^{12} n}\frac{1}{\log^{2} n}\right) \leq 2^{-\frac{n}{\log^{14} n}}.$$

This shows that our protocol is a reducer as claimed. The specifications of this reducer are slightly different from the one in the existence proof (Section 2), but the analysis of Section 3 holds using this new type, and the proof that our construction answers the demands follows.

References

[1] M. Ajtai, and N. Linial: The influence of large coalitions, *Combinatorica*, **13** (1993) 129–145.

[2] N. Alon, and M. Naor: Coin-flipping games immune against linear-sized coalitions, *SIAM J. Computing*, **22** (1993) 403–417.

[3] N. Alon, and J. Spencer: *The Probabilistic Method*, Wiley, 1992.

[4] M. Ben-Or, and N. Linial: Collective coin flipping, in *Randomness and Computation*, (S. Micali, ed.) Academic Press, 1989, 91–115.

[5] A. Cohen, and A. Wigderson: Dispersers, deterministic amplification, and weak random sources, *Proc. 30th IEEE Symp. on the Foundations of Computer Science* (1989) 14–19.

[6] P. Feldman, and S. Micali: Optimal algorithms for Byzantine Agreement, *Proc. 20th Annual ACM Symp. on Theory of Computing* (1988) 148–161.

[7] O. Goldreich, S. Goldwasser, and N. Linial: Fault-tolerant computation in the full information model, *32nd Symposium on the Foundations of Computer Science* (1991) 447–457.

[8] J. Kahn, G. Kalai, and N. Linial: The influence of variables on boolean functions, *29th Symposium on the Foundations of Computer Science*, White Planes, 1988, 68–80.

[9] N. Linial: Games Computers Play – Game-Theoretic Aspects of Computer Science, in: *Handbook of Game Theory with Economic Applications* (R. J. Aumann and S. Hart eds.) North Holland, to appear 1994. Also available as a Leibniz Center, Computer Science Dept. Hebrew University, Tech. Report 5–92.

[10] A. Lubotzky, R. Phillips, and P. Sarnak: Ramanujan Graphs, *Combinatorica* **8** (1988), 261-277.

[11] I. S. Rombauer, and M. Rombauer Becker: *Joy of Cooking*, Hobbs-Merril, 1979, 154.

[12] M. Saks: A robust noncryptographic protocol for collective coin flipping, *SIAM J. Discrete Math.* **2** (1989) 240–244.

Jason Cooper

c/o Nathan Linial

Nathan Linial

Institute of Computer Science
Hebrew University
Jerusalem 91904, Israel
nati@cs.huji.ac.il