Distributed Computing manuscript No.
(will be inserted by the editor)

Dana Angluin · James Aspnes · David Eisenstat

A Simple Population Protocol for Fast Robust Approximate Majority

Received: date / Accepted: date

Abstract We describe and analyze a 3-state one-way population protocol to compute approximate majority in the model in which pairs of agents are drawn uniformly at random to interact. Given an initial configuration of x's, y's and blanks that contains at least one non-blank, the goal is for the agents to reach consensus on one of the values x or y. Additionally, the value chosen should be the majority non-blank initial value, provided it exceeds the minority by a sufficient margin. We prove that with high probability n agents reach consensus in $O(n \log n)$ interactions and the value chosen is the majority provided that its initial margin is at least $\omega(\sqrt{n} \log n)$. This protocol has the additional property of tolerating Byzantine behavior in $o(\sqrt{n})$ of the agents, making it the first known population protocol that tolerates Byzantine agents.

Keywords Population protocols · majority · epidemics · Byzantine faults

1 Introduction

Population protocols [2] model distributed systems in which individual agents are extremely limited, in fact finite-state, and complex behavior of the system as a whole emerges from the rules governing pairwise interaction of the agents. A general survey of results concerning population protocols may be found in [5]; for a detailed comparison with message passing models, see [4]. We describe the model formally in Section 2.

Such models have been defined and used in other fields, for example, statistics, epidemiology, physics and chemistry; understanding their behavior is a fundamental scientific problem. The new perspective we bring as computer scientists is to ask what computational behaviors these systems can exhibit. In addition to fundamental scientific knowledge, answers may provide novel designs for distributed computational systems at many scales.

Chemists have defined a standard model of small molecules in a well-mixed solution, in which the molecules are agents, the state of an agent represents the chemical species of the molecule, and interaction rules specify the probable products of a collision between two molecules; the sequence of collisions is determined by uniform random draws of a pair of agents to interact [9, 10].

In [2] it is shown that this model in principle permits the design of a "computer in a beaker," that is, we can design interaction rules that allow a population of n molecules to simulate the behavior of a register machine with a constant number of registers holding numbers of magnitude $O(n)$ for $poly(n)$ steps with error probability $1/poly(n)$ in parallel time that is a factor of $poly(n)$ larger than the number of simulated instructions. In [3] we have shown that a careful analysis of the properties of epidemics permits us to design a much more efficient simulation, in which the per-step slowdown factor is $O(\log^5 n)$ parallel time.[1] A remaining bottleneck in this construction is the need to perform comparisons between the number of agents in two different states x and y; this is done in [3] using a roundabout algorithm that alternates phases of having x and y tokens cancel each other out upon meeting with phases of doubling the number of x and y tokens in the population. The need for a faster mechanism for computing a majority value was the main driving factor behind the present work.

We present a very simple population protocol (with only 3 states per agent, including the input states x and y) that computes the majority value quickly, provided the initial majority is sufficiently large (Section 3). The essential idea of the protocol is that when two agents with different preferences meet, one drops its preference and enters a special "blank" state b; a blank agent then

Dana Angluin and James Aspnes
Yale University, Department of Computer Science
E-mail: {dana.angluin, james.aspnes}@yale.edu

David Eisenstat
E-mail: eisenstatdavid@gmail.com

[1] Erroneously reported as $O(\log^4 n)$ in [3].

adopts the preference of any non-blank agent it meets. Collisions between agents with opposite preferences reduce support for each preference equally on average. But because a blank agent is more likely meet an agent with the majority preference, encounters between non-blank and blank agents preferentially increase the majority. This creates a strong pressure toward the majority value, and accounts for both the speed and effectiveness of the protocol when the initial majority is sufficiently large. Once all tokens have the same preference, the protocol has converged—further transitions have no effect on the states of the agents.

Unfortunately, while the protocol itself is simple, proving that it converges quickly appears to be very difficult. We design a potential function that approximates the time to convergence from any given state, and show that this potential function converges to its minimum in $O(n \log n)$ interactions with high probability (Section 4). We also show using a separate argument that the output value correctly reports the initial majority with high probability if the net majority is $\omega(\sqrt{n} \log n)$; the essential idea is that we can bound this process by bounding the net majority from below with a coupled fair random walk, and show that there is not enough time for the random walk to reach 0 before convergence if the initial majority is large enough (Section 5).

We then consider some variants of the basic model. In Section 6, we show that correctness continues to hold (with a larger initial net majority) even if agents initially do not participate in the protocol but are recruited upon receiving a start signal that propagates via epidemic. This is needed for the register machine simulation discussed above, because a comparison operation using the approximate majority protocol is triggered in exactly this manner. In Section 7, we consider the effect of including Byzantine agents into the model. These are agents that can pretend to be in any state in an interaction. We show that, with high probability, $o(\sqrt{n})$ Byzantine agents cannot significantly delay convergence of the protocol to a state where most normal agents record the correct majority, although they can keep a small proportion of the normal agents confused, and (after exponential time on average) they can eventually drive the protocol to a stable bad state where all normal agents are blank. Finally, in Section 8, we consider the case where there are more than two possible input values to choose between, and describe a reduction to the two-valued protocol that converges one bit at a time to a common consensus value in $O(kn \log n)$ interactions, where k is the number of bits needed to represent an input symbol.

2 Model

A **population protocol** consists of a finite set of states Q, a finite set of input symbols $X \subseteq Q$, a finite set of output symbols Y, an output function $\gamma : Q \to Y$, and a **joint transition function** $\delta : Q \times Q \to Q \times Q$. A population protocol is executed by a fixed finite **population** of agents with states in Q. For convenience, we assume that each agent has an identity $v \in V$, but agents do not know their own identities or others'.

Initially, each agent is assigned a state according to an **input** $x : V \to X$ that maps agent identities to input symbols. In the general population protocol model, there is an **interaction graph**, a directed graph $G = (V, A)$ without self-loops, whose arcs indicate the possible agent interactions that may take place. (G is directed because we assume that interacting agents are able to break symmetry.) In this paper, G will always be a complete graph.

During each execution step, an arc (v, w) is chosen uniformly at random from A. The "source" agent v is the **initiator**, and the "sink" agent w is the **responder**. These agents update their states jointly according to δ: if v is in state q_v and w is in state q_w, the state of v becomes $\delta_1(q_v, q_w)$, the state of w becomes $\delta_2(q_v, q_w)$, where δ_i gives the i^{th} coordinate of the output of δ. The states of all other agents are unchanged. For any given V, a population protocol **computes** a (possibly partial) function $g : X^V \to Y$ in ℓ steps with error probability ϵ if for all $x \in g^{-1}(Y)$, the configuration $c : V \to Q$ reached after ℓ steps satisfies the following properties with probability $1 - \epsilon$.

- All agents agree on the correct output: for all $v \in V$, $g(x) = \gamma(c(v))$.
- This is also true of every configuration reachable from c.

We are interested in the behavior exhibited by a fixed protocol running in any finite population. Given a family of functions defined for all finite populations (e.g., majority) we ask how well a fixed protocol can compute the function in each finite population.

Although we have described the population protocol model in a sequential light, in which each step is a single pairwise interaction, interactions between pairs involving different agents are independent and may be thought of as occurring in parallel. In measuring the speed of population protocols, then, we define 1 unit of **parallel time** to be $|V|$ steps. The rationale is that in expectation, each agent initiates 1 interaction per parallel time unit; this corresponds to the chemists' idealized assumption of a well-mixed solution.

2.1 Byzantine Agents

We extend the basic randomized population protocol model described above to allow Byzantine behavior from some of the agents. In addition to the n normal agents we allow a population to include z Byzantine agents. For each interaction, an ordered pair of agents is selected uniformly at random from the population of normal and Byzantine agents. A Byzantine agent may simulate any normal agent state in an interaction, and its choice of

state may depend on both the global configuration and the identity of the specific agent it encounters. The state of Byzantine agents is not meaningful and so is not included in the description of a configuration. We first describe our protocol and analyze its behavior without Byzantine agents.

3 A 3-State Approximate Majority Protocol

We analyze the behavior of the following population protocol with states $Q = \{b, x, y\}$. The state b is the **blank** state. Row labels give the initiator's state and column labels the responder's state.

	x	b	y
x	(x,x)	(x,x)	(x,b)
b	(b,x)	(b,b)	(b,y)
y	(y,b)	(y,y)	(y,y)

Note that this protocol is **one-way**: every interaction changes at most the responder's state; thus it can be implemented with one-way communication. Only the interactions xb, yb, xy, and yx change the responder's state; we may think of these as the only interactions that consume energy. The **blank configuration** of all b's is stable, but cannot be reached from any non-blank configuration because no interaction can eliminate the last x or y. The configurations of all x's and all y's are stable, and every non-blank configuration can reach at least one of them.

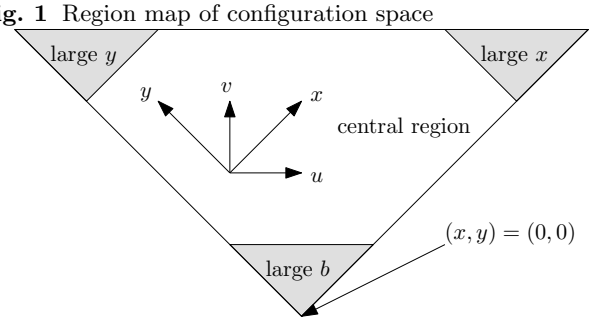
An intuitive description of the process is that agents in state b are undecided, while initiators in states x and y are attempting to convert responders that they meet to adopt their respective states. Such an initiator immediately converts an undecided responder, but only succeeds in reducing an opposing responder to undecided status. The process may also be thought of as two competing epidemics, x's and y's, with the ability to reverse each other's progress.

In Sections 4 and 5, we show that with high probability this protocol (a) converges from any non-blank configuration to a stable configuration in $O(n \log n)$ interactions; and (b) correctly computes the initial majority x or y value provided $\omega(\sqrt{n} \log n)$ more agents carry this value in the starting configuration than carry the opposing value. In Section 7, we show that it can tolerate $o(\sqrt{n})$ Byzantine agents; the formal definition of this property is given there.

4 Convergence

We show that, from any non-blank initial configuration, the 3-state approximate majority protocol converges to either all x tokens or all y tokens within $O(n \log n)$ interactions with high probability. We divide the space of non-blank configurations into four regions: three corners, where most tokens are b, x, or y, and a central region where the tokens are more evenly balanced; see Figure 1. We show that the number of interactions in each region is bounded by $O(n \log n)$ with high probability, by constructing a family of supermartingales of the form $M = e^{aS/n} f(x, y)$ where $a > 0$ is a constant, S counts the number of interactions of a particular type and f is a potential function defined across the entire space of configurations. (We overload x, y and b to denote the number of each token in a configuration.)

Fig. 1 Region map of configuration space



The resulting proof requires a careful selection of f. To keep the argument at least locally simple, we construct separate potential functions to bound different classes of operations, based on the type of interaction that occurs and which region of the configuration space it occurs in. The reason for this classification is that the behavior of the protocol is qualitatively different in different regions of the configuration space. When most tokens are blank, the protocol acts like an epidemic, with non-blank tokens rapidly infecting blank tokens. When most tokens carry the same non-blank value, the protocol acts like coupon collector, with the limit on convergence being the time for the few remaining minority tokens to be converted to the majority value. In the central region, where no token type predominates, the protocol acts like a random walk with increasing bias away from the center. Unfortunately, in none of these configurations does the protocol act *enough* like the analogous well-known stochastic processes to permit a direct reduction to previous results, and the behavior in border areas blends smoothly between one form and another. The supermartingale/potential function approach allows separate arguments designed for each region to be blended smoothly together. Unfortunately, this still requires considerable calculation to verify that each potential function does what it is supposed to.

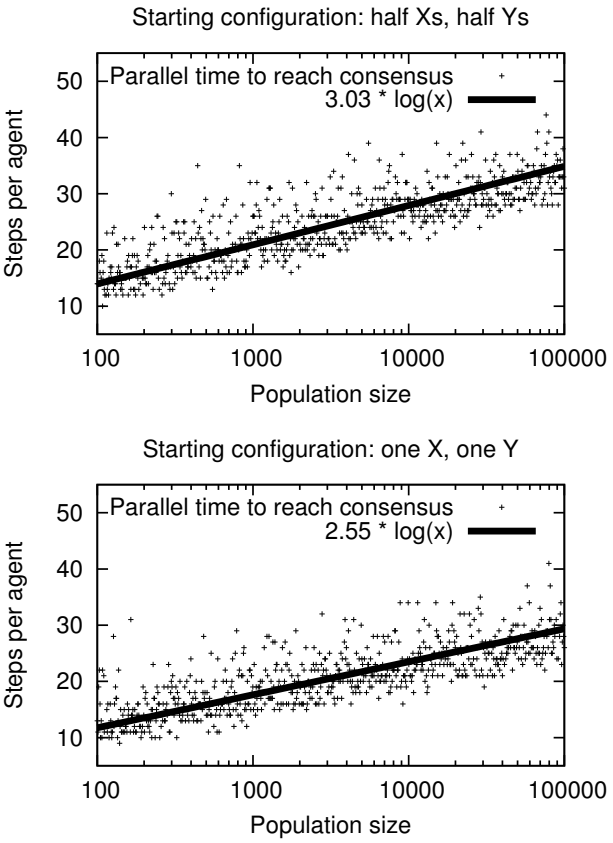
The reader may be surprised to find that such a simple protocol requires such a lengthy proof. Despite substantial efforts, we were unable to apply more powerful tools to this problem. Part of the reason is that we are trying to obtain exact asymptotic bounds on a system in which much of the interesting behavior occurs when particular tokens are very rare or when the behavior of

the protocol is highly random (e.g., with evenly balanced numbers of x and y tokens); this (together with the fact that the corresponding systems of differential equations do not have closed-form solutions) appears to rule out arguments based on classical techniques involving reduction to a continuous process in the limit (e.g., [12, 14]). Similarly, approaches based on direct computation of hitting times or eigenvalues of the resulting Markov chain would appear to require substantially more work than a direct potential function argument.

It is possible that such difficulties are an inherent property of randomized population protocols. The ability to construct register machines using such protocols [2,3] suggests that analysis of an arbitrary protocol for arbitrarily large populations quickly enters the realm of undecidability. For example, the question of whether a given protocol computes the constant function 0 with probability $(1 - 1/n)$ in every possible population is undecidable. But we cannot rule out the possibility that a more sophisticated approach might give an easier proof of the convergence rate for the particular protocols we are interested in.

Our results are stated using explicit constant factors. The reader should be warned that in many cases these are gross overestimates, and that from simulation we observe that the expected number of interactions to convergence seems to be less than $4n \log n$ from two challenging families of initial configurations (see Figure 2.) The first of these, initial populations evenly divided between x and y with no blank tokens, can be shown numerically for reasonably small n to be the configurations that maximize expected convergence time.

Fig. 2 Simulation results: parallel time of approximate majority from two initial conditions



4.1 Notation and Preliminaries

We write x_t, y_t, and b_t for the number of x, y, and blank tokens at time t (that is, following t interactions). When it will not cause confusion, we will omit the subscripts. We are interested in properties of the discrete time stochastic process

$$(x_0, y_0, b_0), (x_1, y_1, b_1), (x_2, y_2, b_2), \ldots$$

giving the values of these quantities after each interaction. Let τ_* denote the **convergence time**, defined to be the first time t at which $x_t = n$ or $y_t = n$, indicating that the agents have reached consensus.

Formally, for each t we consider the σ-algebra generated by $\{(x_i, y_i, b_i)\}$ for all $i \leq t$, which we denote $\mathcal{F}_t$. To avoid writing $\mathcal{F}_t$ everywhere, we will implicitly condition any probabilities or expected values concerning a single interaction ending at some time t on $\mathcal{F}_{t-1}$.

To reduce the size of some of the expressions we will be dealing with, we introduce several variables for referring to frequently-occurring expressions. These are as follows.

$$u = x - y$$

$$v = x + y = n - b$$

$$g = 1/(n(n-1))$$

Note that $-n \leq u \leq n$, and $|u| = n$ indicates that convergence has been reached. Also $0 \leq v \leq n$, with $1 \leq v$ for non-blank configurations. The change of basis to u and v allows us to take advantage of the symmetry between x and y tokens. The variable g is the conversion factor between numbers of pairs of tokens and the probability that one of these pairs is selected; thus, for example, gvb gives the probability of an interaction with a non-blank initiator and a blank responder.

We make extensive use of the Δ operator from the theory of difference equations, defined as $(\Delta f)_t = f_{t+1} - f_t$.

We use 0-1 indicator variables for various events, writing for example I_t^{yb} for the indicator of the event that the interaction that ends at time t is an xb or a yb interaction. Though we attempt to give these indicator variables evocative names, we prefer convenience to absolute consistency: so, for example, we use I^{xy} as the indicator for the event of either an xy or a yx interaction. Table 1 lists the indicator variables we use.

Indicator	Sum	Event
I^{vb}	S^{vb}	xb or yb interaction
I^{xy}	S^{xy}	xy or yx interaction
I^b	S^b	b corner interaction with $b \geq (7/8)n$
I^x	S^x	x corner interaction with $x \geq (7/8)n$
I^y	S^y	y corner interaction with $y \geq (7/8)n$
I^c	S^c	central interaction: $I^b = I^x = I^y = 0$
I^z	S^z	interaction with a Byzantine initiator

Table 1 Indicator variables and their sums.

For each indicator variable I_t we use a corresponding variable $S_t = \sum_{j=1}^{t} I_j$ for the total number of times I's event has occurred.

4.2 More on Probability

In its simplest form, a **supermartingale** is a sequence of real-valued random variables $X_0, X_1, X_2, X_3, \ldots$ where each X_t has bounded expectation and the conditional expectation $\mathrm{E}[X_t | X_0 \ldots X_{t-1}] \leq X_{t-1}$ (see [11, Chapter 12]). The intuition is that a supermartingale is a process that always stays the same or drops on average. More generally, a supermartingale can be equipped with a sequence of σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \ldots$ that express the information available at time t; here the requirements is that each X_t is measurable with respect to $\mathcal{F}_t$, and that $\mathrm{E}[X_t | \mathcal{F}_{t-1}] \leq X_{t-1}$. In either case the expected value of each X_t is bounded by the initial value: $\mathrm{E}[X_t] \leq \mathrm{E}[X_0]$.

Related processes include **submartingales**, where $\mathrm{E}[X_t | X_0 \ldots X_{t-1}] \geq X_{t-1}$, and **martingales**, where $\mathrm{E}[X_t | X_0 \ldots X_{t-1}] = X_{t-1}$ (these are also discussed in [11]). A useful tool for bounding the value of a martingale is Azuma's inequality [1]. This says that if the differences $X_t - X_{t-1}$ in a martingale with $X_0 = 0$ are all bounded by by 1, then $\Pr[X_t > \lambda] \leq \exp(-\lambda^2/2t)$, and by symmetry $\Pr[X_t < -\lambda] \leq \exp(-\lambda^2/2t)$; the proof is by bounding $\mathrm{E}[\exp(\alpha X_t)]$ for a suitable choice of α. The upper bound also holds for supermartingales (as observed in [14]); the intuition is that any extra drop in X_t only makes it harder to exceed λ.

For processes with varying lengths, it is useful to consider the concept of a **stopping time**. This is a random variable τ that is finite with probability 1 and for which the event $[\tau = t]$ can be determined by examining $X_0 \ldots X_t$ (or, in the more general case, $\mathcal{F}_t$). A (super/sub)martingale truncated by a stopping time by setting $X'_t = X_{\min(t,\tau)}$ is still a (super/sub)martingale, and the property $\mathrm{E}[X_\tau] \leq \mathrm{E}[X_0]$ holds for supermartingales [6, 11].

In our case, we let τ_* be the stopping time at which the protocol converges, and let $\tau = \min(\tau_*, kn \log n)$ for some fixed k. Note that τ is also a stopping time. This truncation guarantees that τ and quantities defined in terms of it are finite and well-defined, despite the logical possibility that convergence is not achieved and τ_* is undefined. Assuming that the potential function f does not vary too much over the space of configurations, we can use the supermartingale property $\mathrm{E}[M_\tau] \leq M_0$ to show that $e^{aS_\tau/n}$ is small, and then use Markov's inequality to get the bound on S_τ. Summing the bounds for each region then gives the total bound on the number of interactions. Though it would seem that truncating at time $kn \log n$ assumes what we are trying to prove, in fact we show that with high probability the total number of interactions is much less than $kn \log n$, implying that we do in fact converge by the given time bound.

4.3 Relative Changes in $1/f$

Some of our potential functions are of the form $1/f$ for some f for which it is easy to compute Δf. The following lemma relates the relative change in $1/f$ to the values of Δf and f. We use this to show that various functions $1/f$ drop by a constant factor conditioned on certain events, which will pay for the rise in the exponential factor to which they are attached.

Lemma 1 *Let f be a sequence of positive real numbers such that $|\Delta f/f| < 1$. Then*

$$\begin{aligned}\frac{\Delta(1/f)}{1/f} &= -\Delta f/f + (\Delta f/f)^2 - (\Delta f/f)^3 + \ldots \\ &= \sum_{i=1}^{\infty} (-\Delta f/f)^i. \end{aligned} \tag{1}$$

Proof Compute

$$\begin{aligned}\frac{\Delta(1/f)}{1/f} &= f\Delta(1/f) \\ &= f\left(\frac{1}{f+\Delta f} - \frac{1}{f}\right) \\ &= f\left(\frac{f-(f+\Delta f)}{(f+\Delta f)f}\right) \\ &= \frac{-\Delta f}{f+\Delta f} \\ &= \frac{-\Delta f/f}{1+\Delta f/f} \\ &= (-\Delta f/f)\sum_{i=0}^{\infty}(-\Delta f/f)^i \\ &= \sum_{i=1}^{\infty}(-\Delta f/f)^i. \end{aligned}$$

□

4.4 Bounding the Number of State-Changing Interactions

We start by obtaining a bound on the number of interactions in which some agent changes state: specifically,

the xb, yb, xy, and yx interactions. Define

$$S_t^{vb} = \sum_{i=1}^{t} I_i^{vb},$$

the number of interactions out of the first t that are of type xb or yb and

$$S_t^{xy} = \sum_{i=1}^{t} I_i^{xy},$$

the number that are of type xy or yx. Because every blank created by an xy or yx interaction must be converted back to an x or y by an xb or yb interaction before participating in another xy or yx interaction, for any t we have

$$S_t^{xy} \leq S_t^{vb} + n.$$

At convergence ($t = \tau_*$) there are no blanks and

$$S_{\tau_*}^{xy} \leq S_{\tau_*}^{vb}.$$

We show that the potential function $\frac{1}{u^2+2n}$ is reduced by $-\Theta(1/n)$ of its previous value on average conditioned on the event that either an xb or a yb interaction occurs, and that it rises by a smaller relative amount conditioned on the event that either an xy or a yx interaction occurs. This gives a high-probability $O(n \log n)$ bound on $S_{\tau_*}^{vb} - \alpha S_{\tau_*}^{xy}$ for some constant $0 < \alpha < 1$. But since $S_{\tau_*}^{vb} \geq S_{\tau_*}^{xy}$ we can derive an $O(n \log n)$ bound on $S_{\tau_*}^{vb}$ alone (and thus also on $S_{\tau_*}^{xy}$ alone) from the bound on $S_{\tau_*}^{vb} - \alpha S_{\tau_*}^{xy}$.

Our goal is thus to bound the change in $1/f$ for $f = u^2 + 2n$. We start by getting a bound on $|\Delta f/f|$, which limits how many terms we need to consider in the series expansion in (1). Compute

$$\begin{aligned}\Delta f &= ((u + \Delta u)^2 + 2n) - (u^2 + 2n)\\ &= u^2 + 2u\Delta u + (\Delta u)^2 - u^2\\ &= 2u\Delta u + (\Delta u)^2.\end{aligned}$$

Because $u = x - y$, Δu can only be -1, 0 or 1. Since $|\Delta u| \leq 1$, we have $|\Delta(u^2 + 2n)| \leq 2|u| + 1$, and thus $|\Delta f/f| \leq (2|u| + 1)/(u^2 + 2n) = O(\min(1/|u|, |u|/2n))$. This last quantity is maximized at $u = \Theta(\sqrt{n})$, giving the bound

$$|\Delta f/f| = O(1/\sqrt{n}). \tag{2}$$

We now consider the expected values of Δf and $(\Delta f)^2$ conditioned on I^{vb} or I^{xy}. In computing these expectations, we also condition implicitly on the state before the transition (so, for example, b, x, and y are all treated as constants).

$$\begin{aligned}&\mathrm{E}[\Delta(u^2 + 2n)|I^{vb}]\\ &\quad = (x/v)(2u + 1) + (y/v)(-2u + 1)\\ &\quad = 2u(x - y)/v + (x + y)/v\\ &\quad = 2u^2/v + 1.\end{aligned}$$

$$\begin{aligned}&\mathrm{E}[\Delta(u^2 + 2n)|I^{xy}]\\ &\quad = \frac{1}{2}(2u + 1) + \frac{1}{2}(-2u + 1)\\ &\quad = 1.\end{aligned}$$

$$\begin{aligned}&\mathrm{E}[(\Delta(u^2 + 2n))^2]|I^{vb}]\\ &\quad = (x/v)(2u + 1)^2 + (y/v)(-2u + 1)^2\\ &\quad = 4u^2 + 4u(x - y)/v + 1\\ &\quad = 4u^2 + 4u^2/v + 1.\end{aligned}$$

$$\begin{aligned}&\mathrm{E}[(\Delta(u^2 + 2n))^2|I^{xy}]\\ &\quad = \frac{1}{2}(2u + 1)^2 + \frac{1}{2}(-2u + 1)^2\\ &\quad = 4u^2 + 1.\end{aligned}$$

Applying Lemma 1 to the relative change in $1/(u^2 + 2n)$ conditioned on I^{vb} gives the following.

Lemma 2

$$\mathrm{E}\left[\frac{\Delta(1/(u^2 + 2n))}{1/(u^2 + 2n)}|I^{vb}\right] \leq -\frac{15}{32}n^{-1} + O(n^{-3/2}). \tag{3}$$

Proof Let $r = u^2/n$, so that $u^2 = rn$. Let $f = u^2 + 2n$. Then by (2), $|f/\Delta f| = O(n^{-1/2})$ and we apply Lemma 1 to get the following.

$$\begin{aligned}&\mathrm{E}\left[\frac{\Delta(1/f)}{1/f}|I^{vb}\right]\\ &= \mathrm{E}\left[-\frac{\Delta f}{f} + \left(\frac{\Delta f}{f}\right)^2 + O(n^{-3/2})|I^{vb}\right]\\ &= -\frac{2u^2/v + 1}{u^2 + 2n} + \frac{4u^2 + 4u^2/v + 1}{(u^2 + 2n)^2} + O(n^{-3/2})\\ &= \frac{-(2u^2/v + 1)(u^2 + 2n) + 4u^2 + 4u^2/v + 1}{(u^2 + 2n)^2}\\ &\qquad + O(n^{-3/2})\\ &= \frac{-2u^4/v - 4u^2n/v - u^2 - 2n + 4u^2 + 4u^2/v + 1}{(u^2 + 2n)^2}\\ &\qquad + O(n^{-3/2})\\ &= \frac{(-2u^4 - 4u^2n + 4u^2)/v - 2n + 3u^2 + 1}{(u^2 + 2n)^2} + O(n^{-3/2})\\ &\leq \frac{(-2u^4 - 4u^2n + 4u^2)/n - 2n + 3u^2 + 1}{(u^2 + 2n)^2} + O(n^{-3/2})\end{aligned}$$

(because $-2u^4 - 4u^2n + 4u^2 \leq 0$ and $v \leq n$)

$$= \frac{(-2r^2n^2 - 4rn^2 + 4rn)/n - 2n + 3rn + 1}{(r + 2)^2n^2} + O(n^{-3/2})$$

$$= \frac{-2r^2n - rn - 2n + 4r + 1}{(r+2)^2n^2} + O(n^{-3/2})$$
$$= (1/n)\frac{-2r^2 - r - 2 + 4r/n + 1/n}{(r+2)^2} + O(n^{-3/2})$$
$$\leq (1/n)\frac{-2r^2 - r - 2}{(r+2)^2} + 1/n^2 + O(n^{-3/2}),$$

(because $(4r+1)/(r+2)^2 \leq 1$)

$$= (1/n)\frac{-2r^2 - r - 2}{(r+2)^2} + O(n^{-3/2}).$$

Some quick differentiation reveals that $(-2r^2 - r - 2)/(r+2)^2$ takes on its maximum value of $-15/32$ at $r = 2/7$. The claimed bound follows. □

Conditioning on I^{xy}, it is possible for $1/(u^2+2n)$ to rise slightly. But we can again bound the rise using Lemma 1.

Lemma 3

$$\mathrm{E}\left[\frac{\Delta(1/(u^2+2n))}{1/(u^2+2n)}|I^{xy}\right] \leq \frac{9}{32}n^{-1} + O(n^{-3/2}). \tag{4}$$

Proof Again let $r = u^2/n$ and $f = u^2 + 2n$. Now we have

$$\mathrm{E}\left[\frac{\Delta(1/f)}{1/f}|I^{xy}\right]$$
$$= \mathrm{E}\left[-\frac{\Delta f}{f} + \left(\frac{\Delta f}{f}\right)^2 + O(n^{-3/2})|I^{xy}\right]$$
$$= -\frac{1}{u^2+2n} + \frac{4u^2+1}{(u^2+2n)^2} + O(n^{-3/2})$$
$$= \frac{-u^2 - 2n + 4u^2 + 1}{(u^2+2n)^2} + O(n^{-3/2})$$
$$= \frac{3u^2 - 2n + 1}{(u^2+2n)^2} + O(n^{-3/2})$$
$$= \frac{3rn - 2n + 1}{(r+2)^2n^2} + O(n^{-3/2})$$
$$= (1/n)\frac{3r - 2 + 1/n}{(r+2)^2} + O(n^{-3/2})$$
$$= (1/n)\frac{3r-2}{(r+2)^2} + O(n^{-3/2}).$$

Here we obtain a maximum for $(3r-2)/(r+2)^2$ of $9/32$ at $r = 10/3$. □

Shifting both coefficients up slightly, from $-15/32$ to $-7/16$ and from $9/32$ to $5/16$, we have the following.

Corollary 1 *For all sufficiently large n,*

$$\mathrm{E}[1/f_{t+1}|I^{vb}] < \exp(-\frac{7}{16}n^{-1})(1/f_t)$$

and

$$\mathrm{E}[1/f_{t+1}|I^{xy}] < \exp(\frac{5}{16}n^{-1})(1/f_t).$$

Proof From Lemma 2,

$$\mathrm{E}[\Delta(1/f)/(1/f)|I^{vb}] \leq -\frac{15}{32}n^{-1} + O(n^{-3/2}),$$

which implies

$$\mathrm{E}[((1/f_{t+1}) - (1/f_t))/(1/f_t)|I^{vb}] \leq -\frac{15}{32}n^{-1} + O(n^{-3/2}),$$

and therefore

$$\mathrm{E}[1/f_{t+1}|I^{vb}] \leq (1 - \frac{15}{32}n^{-1} + O(n^{-3/2}))(1/f_t).$$

For all sufficiently large n,

$$(1 - \frac{15}{32}n^{-1} + O(n^{-3/2})) < \exp(-\frac{7}{16}n^{-1}),$$

because shifting $-15/32$ up to $-7/16$ absorbs both the $O(n^{-3/2})$ error term and the second-order and higher terms in the Taylor series expansion of the exponential. The second claim follows similarly from Lemma 3 by shifting $9/32$ up to $5/16$. □

For other interactions the value of $1/f$ is unchanged. Combining these gives the following.

Lemma 4 *The stochastic process $\{M_t\}$ given by*

$$M_t = \frac{\exp\left(\left(\frac{7}{16}S_t^{vb} - \frac{5}{16}S_t^{xy}\right)/n\right)}{u_t^2 + 2n} \tag{5}$$

is a supermartingale.

Proof We analyze the expected value of M_t given the history of the process through interaction $t-1$; it suffices to show that conditioning on the indicator I for each possible type of interaction, $\mathrm{E}[M_t|I] \leq M_{t-1}$. If the interaction is not of type vb or xy, then $M_t = M_{t-1}$. If the interaction is of type vb then $S_t^{vb} = 1 + S_{t-1}^{vb}$ and $\mathrm{E}[1/f_t|I^{vb}] < \exp(-\frac{7}{16}n^{-1})(1/f_{t-1})$, so $\mathrm{E}[M_t|I^{vb}] < M_{t-1}$. Finally, if the interaction is of type xy then $S_t^{xy} = 1 + S_{t-1}^{xy}$ and $\mathrm{E}[1/f_t|I^{xy}] < \exp(\frac{5}{16}n^{-1})(1/f_{t-1})$, so $\mathrm{E}[M_t|I^{xy}] < M_{t-1}$. □

Corollary 2 *Let $\tau \leq \min(\tau_*, M)$ be a stopping time, where τ_* is the time at which $x = n$ or $y = n$ first holds and M is an arbitrary bound. Then for all sufficiently large n,*

$$\Pr\left[S_\tau^{vb} \geq 8n\log(n+2) + 8cn\log n + (5/2)n\right] \leq n^{-c}. \tag{6}$$

Proof Since $\{M_t\}$ as defined in Lemma 4 is a supermartingale, we have $\mathrm{E}[M_\tau] \leq M_0 \leq 1/n$. Since $|u| \leq n$ at τ, the denominator in M_τ is at most $n^2 + 2n$, giving the bound

$$\mathrm{E}\left[\exp\left(\left(\frac{7}{16}S_\tau^{vb} - \frac{5}{16}S_\tau^{xy}\right)/n\right)\right] \leq \frac{n^2+2n}{n} = n + 2.$$

Markov's inequality now gives

$$\Pr\left[\exp\left(\left(\frac{7}{16}S_\tau^{vb} - \frac{5}{16}S_\tau^{xy}\right)/n\right) \geq \alpha(n+2)\right] \leq \frac{1}{\alpha},$$

from which it follows that

$$\Pr\left[\left(\frac{7}{16}S_\tau^{vb} - \frac{5}{16}S_\tau^{xy}\right) \geq n\log(n+2) + n\log\alpha\right] \leq \frac{1}{\alpha}$$

or, letting $\alpha = n^c$,

$$\Pr\left[\left(\frac{7}{16}S_\tau^{vb} - \frac{5}{16}S_\tau^{xy}\right) \geq n\log(n+2) + cn\log n\right] \leq n^{-c}. \tag{7}$$

Because for all t, $S_t^{xy} \leq S_t^{vb} + n$, we have

$$\frac{7}{16}S_\tau^{vb} - \frac{5}{16}S_\tau^{xy} \geq \frac{1}{8}S_\tau^{vb} - \frac{5}{16}n.$$

Substituting $\frac{1}{8}S_\tau^{vb} - (\frac{5}{16})n$ into (7) gives the result for S_τ^{vb}. Because $S_\tau^{xy} \leq S_\tau^{vb} + n$, we simultaneously bound S_τ^{xy}. □

Because M is arbitrary in this result, we may choose $M > 8n\log(n+2) + 8cn\log n + (5/2)n$. Then if $S_{\tau_*}^{vb}$ exceeds $8n\log(n+2)+8cn\log n+(5/2)n$ so does S_τ^{vb}. Thus we get a high probability $O(n\log n)$ bound on $S_{\tau_*}^{vb}$ and on $S_{\tau_*}^{xy}$, giving a bound on the total number of interactions that change the state of some agent until convergence, i.e., on the total energy cost of the protocol.

In the following sections, we extend this $O(n\log n)$ bound on total energy to an $O(n\log n)$ bound on total interactions, by adding in the non-state changing bb, xx, yy, bx, and by interactions.

4.5 Bounding Interactions in the Central Region

We now consider interactions occurring in configurations with $\max(x,y,b) < \frac{7}{8}n$. The essential idea is that in the central region where this condition holds, at least two of b, x, and y must be at least $n/16$. It follows that whenever an interaction in this region occurs, either $\Pr[I^{xy}=1] \geq 2g(n/16)^2 > 1/128$ (if both x and y are at least $n/16$) or $\Pr[I^{vb}=1] \geq g(n/16)^2 > 1/256$ (if b and either x or y are at least $n/16$.) Recall that $g = 1/n(n-1)$.

Lemma 5 *Let I_t^c be the indicator variable for the event that the interaction at time t starts in a state with $\max(x,y,b) < \frac{7}{8}n$ and let $S_t^c = \sum_{i=1}^t I_i^c$. Let $\tau \leq \min(\tau_*, M)$ be a stopping time, where τ_* is the time at which $x = n$ or $y = n$ first holds and M is an arbitrary bound. Then for all sufficiently large n*

$$\Pr\left[S_\tau^c \geq 130S_\tau^{vb} + 258S_\tau^{xy} + cn\log n\right] \leq n^{-c}. \tag{8}$$

Proof Let

$$C_t = \exp\left(n^{-1}\left(S_t^c - 130S_t^{xy} - 258S_t^{vb}\right)\right).$$

Observe that conditioned on $I_t^c = 1$, if x_{t-1} and y_{t-1} are both at least $n/16$, then there is a probability of at most $(1-1/128)$ of an interaction which increases S_{t-1}^c but not S_{t-1}^{vb} or S_{t-1}^{xy} and multiplies C_{t-1} by $\exp(1/n)$, and a probability of at least $1/128$ of an xy interaction, which increases both S_{t-1}^c and S_{t-1}^{xy} and multiplies C_{t-1} by $\exp(-129/n)$. Similarly, if b and either x or y is at least $n/16$, there is a probability of at most $(1-1/256)$ of an interaction which increases S_{t-1}^c but not S_{t-1}^{vb} or S_{t-1}^{xy} and multiplies C_{t-1} by $\exp(1/n)$, and a probability of at least $1/256$ of a vb interaction which increases both S_{t-1}^c and S_{t-1}^{vb} and multiplies C_{t-1} by $\exp(-257/n)$. Taking the maximum of these two cases, we get the following upper bound.

$$\begin{aligned}
&\mathrm{E}[C_t|\mathcal{F}_{t-1}, I_t^c = 1] \\
&\quad\leq C_{t-1}\max\left\{\begin{array}{l}\left(1-\frac{1}{128}\right)e^{1/n} + \frac{1}{128}e^{-129/n}, \\ \left(1-\frac{1}{256}\right)e^{1/n} + \frac{1}{256}e^{-257/n}\end{array}\right\} \\
&\quad\leq C_{t-1}\max\left\{\begin{array}{l}1 + (1-129/128)n^{-1} + O(n^{-2}), \\ 1 + (1-257/256)n^{-1} + O(n^{-2})\end{array}\right\} \\
&\quad\leq C_{t-1}\max\left\{\begin{array}{l}1 - 1/(128n) + O(n^{-2}), \\ 1 - 1/(256n) + O(n^{-2})\end{array}\right\} \\
&\quad\leq C_{t-1},
\end{aligned}$$

for sufficiently large n. When $I_t^c = 0$, C_t cannot increase, so we have $\mathrm{E}[C_t|\mathcal{F}_{t-1}] \leq C_{t-1}$ always, and C_t is a supermartingale.

It follows that $\mathrm{E}[C_\tau] \leq C_0 = 1$. Applying Markov's inequality as in the proof of Corollary 2 gives

$$\Pr\left[S_\tau^c \geq 130S_\tau^{vb} + 258S_\tau^{xy} + cn\log n\right] \leq n^{-c}. \tag{9}$$

□

4.6 Bounding Interactions with Large b

We now consider interactions in the corner of the space where b is large. Here we use the potential function $f = 1/v$, which drops consistently by $\Theta(-1/n)$ of its current value on average when $b \geq (7/8)n$, and whose rise in other parts of the space is bounded by $O(I^{xy}/n)$. This gives the following result:

Lemma 6 *Let I_t^b be the indicator for the event that the interaction ending at time t starts in a state with $b \geq (7/8)n$. Let $S_t^b = \sum_{i=1}^t I_t^b$. Let $\tau \leq \min(\tau_*, M)$ be a stopping time, where τ_* is the time at which $x = n$ or*

$y = n$ first holds and M is an arbitrary bound. Then for all sufficiently large n

$$\Pr\left[S_\tau^b \geq 4cn\log n + 40S_\tau^{xy}\right] \leq n^{-c}.$$

Proof We first consider the case of just one nonblank, that is $v = 1$. In this case

$$\frac{\Delta(1/v)}{1/v} = -\frac{I^{vb}}{v+1},$$

and

$$\frac{\mathrm{E}[\Delta(1/v)]}{1/v} = -\frac{gvb}{v+1} = -\frac{1}{2n}.$$

Assume $v \geq 2$. The function $1/v$ is simple enough that we can compute $\Delta(1/v)/(1/v)$ directly:

$$\begin{aligned}\frac{\Delta(1/v)}{1/v} &= v\left(I^{vb}\left(\frac{1}{v+1} - \frac{1}{v}\right) + I^{xy}\left(\frac{1}{v-1} - \frac{1}{v}\right)\right)\\ &= v\left(I^{vb}\frac{-1}{v(v+1)} + I^{xy}\frac{1}{v(v-1)}\right)\\ &= -\frac{I^{vb}}{v+1} + \frac{I^{xy}}{v-1}.\end{aligned}$$

We now consider two cases, depending on whether b is larger or smaller than $(7/8)n$. When $b \geq (7/8)n$, we have

$$\begin{aligned}\frac{\mathrm{E}[\Delta(1/v)]}{1/v} &= \mathrm{E}\left[-\frac{I^{vb}}{v+1} + \frac{I^{xy}}{v-1}\right]\\ &= -\frac{gvb}{v+1} + \frac{2gxy}{v-1}\\ &\leq -gb/2 + \frac{2g(v/2)^2}{v-1}\\ &\leq -\frac{(7/8)n}{2n(n-1)} + gv\\ &\text{(because } v/(v-1) \leq 2 \text{ when } v \geq 2)\\ &\leq -(7/16)n^{-1} + \frac{n/8}{n(n-1)}\\ &\text{(because } v \leq n/8)\\ &\leq -(7/16)n^{-1} + (1/8)n^{-1} + O(n^{-2})\\ &= -(5/16)n^{-1} + O(n^{-2}).\end{aligned}$$

Alternatively, when $b < (7/8)n$, we have

$$\begin{aligned}\frac{\Delta(1/v)}{1/v} &= -\frac{I^{vb}}{v+1} + \frac{I^{xy}}{v-1}\\ &\leq \frac{I^{xy}}{n/8-1}\\ &\leq (9/n)I^{xy},\end{aligned}$$

for sufficiently large n. As in Corollary 1, we adjust the coefficients to absorb higher-order terms: $-5/16$ to $-1/4$ and 9 to 10. Then we have that

$$B_t = \frac{\exp\left(n^{-1}\left[\frac{1}{4}\left(\sum_{i=1}^t I_i^b\right) - 10\left(\sum_{i=1}^t I_i^{xy}(1-I_i^b)\right)\right]\right)}{v_t}$$

is a supermartingale.

Applying the supermartingale property to B_τ gives

$$\begin{aligned}&\mathrm{E}\left[B_\tau\right]\\ &\leq \mathrm{E}\left[\frac{\exp\left(n^{-1}\left[\frac{1}{4}\left(\sum_{i=1}^\tau I_i^b\right) - 10\left(\sum_{i=1}^\tau I_i^{xy}(1-I_i^b)\right)\right]\right)}{n}\right]\\ &\leq B_0 \leq 1.\end{aligned}$$

Applying Markov's inequality gives

$$\begin{aligned}&\Pr\left[n^{-1}\left[\frac{1}{4}\left(\sum_{i=1}^\tau I_i^b\right) - 10\left(\sum_{i=1}^\tau I_i^{xy}(1-I_i^b)\right)\right] \geq c\log n\right]\\ &\leq n^{-c}.\end{aligned}$$

Rearranging and observing that $\sum_{i=1}^\tau I_i^{xy}(1-I_i^b) \leq S_\tau^{xy}$, we have

$$\Pr\left[S_\tau^b \geq 4cn\log n + 40S_\tau^{xy}\right] \leq n^{-c}$$

as claimed. □

4.7 Bounding Interactions with Large x or y

For $x \geq (7/8)n$ or $y \geq (7/8)n$ we use the potential functions $3y + b + 1$ or $3x + b + 1$, respectively. As with the large-b case, we bound the total number of steps taken when x or y is large by showing the potential function drops by a factor of $\exp(-\Theta(1/n))$ in these corners and rises by an amount we can bound using previous bounds on S^{vb} and S^{xy}.

Lemma 7 *Let I_t^x be the indicator for the event that the interaction ending at time t starts in a state with $x \geq (7/8)n$. Let $S_t^x = \sum_{i=1}^t I_t^x$. Let $\tau \leq \min(\tau_*, M)$ be a stopping time, where τ_* is the time at which $x = n$ or $y = n$ first holds and M is an arbitrary bound. Then for all sufficiently large n*

$$\begin{aligned}&\Pr\left[S_\tau^x \geq 8n\log(3n+1) + 8cn\log n + 136S_\tau^{vb} + 72S_\tau^{xy}\right]\\ &\leq n^{-c}.\end{aligned}$$

By symmetry the same bound holds for S_τ^y.

Proof We consider the probabilities and effects of interaction types that change the value of $3y + b + 1$. For example, a yb interaction happens with probability gyb and increases y by 1 and decreases b by 1, for a net change of $+2$ to $3y + b + 1$. The analyses of yx, xb and xy interactions proceed similarly.

Suppose $(7/8)n \le x < n$. Then we have

$$\begin{aligned}
&\frac{\mathrm{E}\left[\Delta(3y+b+1)\right]}{3y+b+1} \\
&= g\frac{+2yb + yx - xb - 2xy}{3y+b+1} \\
&= g\frac{+2yb - xb - xy}{3y+b+1} \\
&= g\frac{2yb}{3y+b+1} - g\frac{x(y+b)}{3y+b+1} \\
&\le g\frac{2yb}{3y+b+1} - gx/4
\end{aligned}$$

(because $x < n$ and therefore $(y+b)/(3y+b+1) \ge 1/4$)

$$\le g\frac{n}{16} - (7/32)n^{-1}$$

(because $y+b \le n/8$ and therefore $32yb \le (3y+b+1)n$)

$$\begin{aligned}
&= (1/16)n^{-1} - (7/32)n^{-1} + O(n^{-2}) \\
&= -(5/32)n^{-1} + O(n^{-2}).
\end{aligned}$$

Alternatively, if $x < (7/8)n$, we have $3y + b + 1 > y + b > n/8$. This gives

$$\begin{aligned}
\frac{\Delta(3y+b+1)}{3y+b+1} &\le \frac{2I^{vb} + I^{xy}}{3y+b+1} \\
&\le \frac{2I^{vb} + I^{xy}}{n/8} \\
&= (16I^{vb} + 8I^{xy})n^{-1}.
\end{aligned}$$

Shifting the coefficients up as in Corollary 1 and using an argument similar to that in the proof of Lemma 6 shows that the stochastic process $\{X_t\}$ where

$$X_t = \exp\left(n^{-1}\begin{bmatrix} \frac{1}{8}\left(\sum_{i=1}^{t} I_i^x\right) \\ -\left(\sum_{i=1}^{t}(17I_i^{vb} + 9I_i^{xy})(1 - I_i^b)\right)\end{bmatrix}\right) \cdot (3y_t + b_t + 1)$$

is a supermartingale.

The supermartingale property gives that, for bounded $\tau \le \tau_*$,

$$\mathrm{E}[X_\tau] \le X_0 \le 3n + 1.$$

So by Markov's inequality, we have

$$\Pr\left[\begin{array}{l}\exp\left(n^{-1}\begin{bmatrix}\frac{1}{8}S_\tau^x \\ -(17S_\tau^{vb} + 9S_\tau^{xy})\end{bmatrix}\right) \\ \quad \cdot (3y_\tau + b_\tau + 1) \\ \ge (3n+1)n^c\end{array}\right] \le n^{-c}.$$

The quantity $3y_\tau + b_\tau + 1$ is at least 1 (the case of all x tokens). Substituting 1 for $3y_\tau + b_\tau + 1$ and taking logarithms gives

$$\Pr\left[\begin{array}{l} n^{-1}\left(\frac{1}{8}S_\tau^x - (17S_\tau^{vb} + 9S_\tau^{xy})\right) \\ \ge \log(3n+1) + c\log n\end{array}\right] \le n^{-c}.$$

After further rearrangement this becomes

$$\begin{aligned}
&\Pr\left[S_\tau^x \ge 8n\log(3n+1) + 8cn\log n + 136S_\tau^{vb} + 72S_\tau^{xy}\right] \\
&\quad \le n^{-c}.
\end{aligned}$$

□

The same bound clearly holds for the analogous quantity S_τ^y by symmetry.

4.8 Bounding Total Interactions

Now we can put the results of Corollary 2 and Lemmas 5, 6, and 7 together to obtain a high-probability bound on τ_*, the total number of interactions before convergence. The explicit constants in this theorem are quite large; recall that the simulation results in Figure 2 suggest that the true coefficient of $n \log n$ is less than 4.

Theorem 1 *Let τ_* be the time at which $x = n$ or $y = n$ first holds. Then for any fixed $c > 0$ and sufficiently large n,*

$$\Pr\left[\tau_* \ge 6769n\log n + 6773cn\log n + 2552n\right] \le 5n^{-c}.$$

Proof Let the error parameter $c > 0$ be fixed. Let $\tau = \min(\tau_*, 10^4(c+1)n\log n)$. Observe that $\tau = S_\tau^c + S_\tau^b + S_\tau^x + S_\tau^y$, since every interaction takes place either in the central region or in one of the three corners. Using Lemma 5 for the first term, Lemma 6 for the second term, and Lemma 7 for the third and fourth terms, we first get bounds in terms of S_τ^{vb} and S_τ^{xy}. Excluding an error probability of at most $4n^{-c}$, we have

$$\begin{aligned}
\tau &= S_\tau^c + S_\tau^b + S_\tau^x + S_\tau^y \\
&< \begin{pmatrix} 130S_\tau^{vb} + 258S_\tau^{xy} + cn\log n \\ + \quad 40S_\tau^{xy} + 4cn\log n \\ + 136S_\tau^{vb} + 72S_\tau^{xy} + 8cn\log n + 8n\log(3n+1) \\ + 136S_\tau^{vb} + 72S_\tau^{xy} + 8cn\log n + 8n\log(3n+1)\end{pmatrix}
\end{aligned}$$

$$= 402S_\tau^{vb} + 442S_\tau^{xy} + 21cn\log n + 16n\log(3n+1).$$

By Corollary 2, with error probability n^{-c} we have

$$S_\tau^{vb} < 8n\log(n+2) + 8cn\log n + (5/2)n$$

and therefore also

$$S_\tau^{xy} < 8n\log(n+2) + 8cn\log n + (7/2)n$$

. Thus, with total error probability $5n^{-c}$ we have

$$\begin{aligned}\tau &< 6752n\log(n+2) + 6773cn\log n \\ &\quad + 16n\log(3n+1) + 2552n \\ &< 6769n\log n + 6773cn\log n + 2552n,\end{aligned}$$

for sufficiently large n. But if $\tau < 10^4(c+1)n\log n$, then $\tau_* = \tau$, giving the claimed bound. □

5 Correctness of Approximate Majority

Not only does the 3-state protocol converge quickly, but it also converges to the dominant non-blank value in its input if there is a large enough initial majority.

Theorem 2 *With high probability, the 3-state approximate majority protocol converges to the initial majority value if the difference between the initial majority and initial minority populations is $\omega(\sqrt{n}\log n)$.*

Proof Without loss of generality, assume that the initial majority value is x. We consider a coupled process (u_t, u'_t) where $u_t = (x_t - y_t)$ and u'_t is the sum of a series of fair ± 1 coin flips. Initially $u'_0 = u_0$. Later values of u'_t are specified by giving a joint distribution on $(\Delta u, \Delta u')$. We do so as follows. Let p be the probability that $\Delta u = 1$ and q the probability that $\Delta u = -1$. Then let

$$(\Delta u, \Delta u') = \begin{cases} (0,0) & \text{with probability } 1-p-q, \\ (1,1) & \text{with probability } \frac{1}{2}(p+q), \\ (1,-1) & \text{with probability } p - \frac{1}{2}(p+q), \\ (-1,-1) & \text{with probability } q. \end{cases}$$

The probability in the third case is non-negative if $p/(p+q) = \Pr[\Delta u = 1 | \Delta u \neq 0] \geq \frac{1}{2}$. This holds as long as $u \geq 0$; should u ever drop to zero, we end the process.

Observe that unless this event happens, we have $u_t \geq u'_t$. We can also verify by summing the cases that Δu rises with probability exactly p and drops with probability exactly q; and that $\Delta u'$ rises or drops with equal probability $\frac{1}{2}(p+q)$. So we have $\mathrm{E}[\Delta u'] = 0$ and that $|\Delta u'| \leq 1$, the preconditions for Azuma's inequality.

Theorem 1 shows that the process converges before $O(n\log n)$ interactions with high probability. Suppose the process converges at some time $\tau = O(n\log n)$. Then by Azuma's inequality we have that $|u'_\tau - u'_0| = O(\sqrt{n}\log n)$ throughout this interval with high probability. So if $u'_0 = u_0 = \omega(\sqrt{n}\log n)$, it follows that $u_0 \geq u'_0 \geq 0$ throughout the execution, and in particular that the process does not terminate before convergence and that u is non-negative at convergence. But this excludes the $y = n$ case, so the process converges to the initial majority value. □

6 Correctness with an Epidemic-Triggered Start

In this section we analyze a variant of the 3-state protocol, the **epidemic-triggered approximate majority protocol**, where agents are recruited into the computation by an epidemic. This is important to the application of the 3-state majority protocol to the register machine simulation, as the signal to start the next operation is broadcast from the leader via an epidemic.

In this protocol, in addition to the $b/x/y$ value, each agent has an active/inactive bit. Active agents interact as before, inactive agents become active when an active agent initiates an interaction with them, and all other interactions have no effect. By previous results on epidemics in this model, with high probability, a starting configuration with at least one active agent will reach a configuration where all agents are active within $O(n\log n)$ interactions [3]. For the following theorem we require a somewhat larger initial majority to guarantee convergence to the correct value.

Theorem 3 *Let $\epsilon > 0$. If the difference between the initial majority and initial minority populations is $\Omega(n^{3/4+\epsilon})$ and there is exactly one active agent, then with high probability, the epidemic-triggered approximate majority protocol converges to the initial majority value.*

Proof Without loss of generality, assume x is the majority value. Let the **first half** be the prefix of the execution where there are at most $n^{3/4}$ active agents and the **second half** be the rest. With high probability, the first half is over in $O(n\log n)$ interactions. Let x_{active} be the number of active x agents and y_{active} be the number of active y agents. Let $u_{active} = x_{active} - y_{active}$. We show that with high probability, $u_{active} > 0$ throughout the second half, which establishes the theorem statement in conjunction with Theorem 1.

There are two types of events that can change u_{active}: a non-blank agent becomes active, or there is an interaction between active agents. Let $u_{initial}$ be the number of active initially-x agents minus the number of active initially-y agents. Then $\Delta u_{initial}$ is the change attributable to events of the first type and $\Delta(u_{active} - u_{initial})$ is the change attributable to the second type. During the first half, events of the first type predominate. There are $n^{3/4} - 1$ agents that become active, but the probability of an active-active interaction is at most $(n^{3/4}/n)^2 = n^{-1/2}$. By Azuma's inequality, the number

of active-active interactions is $O(n^{1/2}\log n)$ with high probability.

In order to apply the techniques of Theorem 2, we must establish that x's constitute a majority of the active agents at the end of the first half. We apply Azuma's inequality again, finding that with high probability, the value of $u_{initial}$ is within $O(n^{3/8}(\log n)^{1/2})$ of its expectation throughout the execution. In the second half, this expectation is $\Omega(n^{3/4+\epsilon}(n^{3/4}/n)) = \Omega(n^{1/2+\epsilon})$. We conclude that x's enjoy an advantage of $\Omega(n^{1/2+\epsilon})$ to start the second half with high probability. At this point we recapitulate the analysis from the previous theorem, giving the random walk a $O(n^{1/2}\log n)$ head start, since each active-active interaction from the first half might have increased u_{active} by 2. □

7 Tolerating Byzantine Agents

In this section, we show that the 3-state approximate majority protocol can tolerate z Byzantine agents, where $z = o(\sqrt{n})$, computing the correct majority value in $O(n\log n)$ time with high probability despite their interference. However, to do so we must both assume a somewhat larger initial majority, and slightly relax the criterion for convergence.

The issue with convergence is that Byzantine agents can always pull the normal agents out of a converged configuration. For example, if all normal agents are in the x state, any encounter with a Byzantine initiator can shift the normal agent to a b state, and a second encounter can shift it to a y state, even though there are no normal y agents in the population. So we must accept a small number of normal agents that do not have the correct value.

But in fact the situation is worse: if we run long enough, there exists a trajectory with nonzero probability that takes us to the blank configuration, which is stable. So we must also accept a small probability that we reach the blank configuration quickly, and the assurance that we reach it with probability 1 after a very long time. However, we can show that with high probability neither outcome occurs within a polynomial number of steps.

Our technique is to adjust the potential functions used by the non-Byzantine process to account for Byzantine transitions. We then use these adjusted potential functions to show that (a) strong pressure exists to keep the process out of the large-b corner and in the large-x and large-y corners, and (b) the number of interactions (including Byzantine interactions) to reach the x or y corner is still small.

7.1 Biased-Walk Barriers

Let us begin by showing that it is difficult even for Byzantine agents to force the protocol into a configuration with a low value of $v_t = x_t + y_t$.

Observe that if the Byzantine agents attempt to minimize v, v nonetheless increases at each interaction with likelihood proportional to vb and decreases with likelihood proportional to $2xy + zv$. So the probability of an increase conditioned on any change in v is $vb/(vb+2xy+zv) \geq vb/(vb+v^2/2+zv) = b/(b+z+v/2) \geq b/n$ provided $z \leq v/2$. For large b and small z this gives a random-walk behavior that is strongly biased upwards.

Suppose $\sqrt{n} \leq v \leq n/8$. Then $b \geq (7/8)n$ and $z = o(\sqrt{n}) \ll v/2$, so $\Pr[\Delta v = 1 | \Delta v \neq 0] \geq 7/8$. We wish to bound the probability starting from some initial v_0 in this range that v reaches $\sqrt{n}$ before it reaches $n/8$. Though the probability that v rises or falls changes over the interval, the position of v can be lower-bounded by the position of a coupled variable v' that moves according to a biased random walk with fixed probability $p = 7/8$ of increasing by 1 and $q = 1/8$ of decreasing by 1.

Formally, let $i_1, i_2, \ldots$ be the sequence of times for which $v_{i_j} \neq v_{i_j-1}$. Define v'_j by the rule $v'_0 = v_0$ and

1. $v'_j = v'_{j-1} - 1$ if $v_{i_j} = v_{i_j-1} - 1$, or with probability $\frac{1/8 - \Pr[v_{i_j} = v_{i_j-1} - 1 | v_{i_j-1}]}{\Pr[v_{i_j} = v_{i_j-1}+1 | v_{i_j-1}]}$ if $v_{i_j} = v_{i_j-1} + 1$.
2. $v'_j = v'_{j-1} + 1$ with probability 7/8.

Observe that the probability of the first event, conditioned on $v_{i_j} - 1$, is

$$\begin{aligned}
&\Pr[v_{i_j} = v_{i_j-1} - 1] \\
&\quad + \Pr[v_{i_j} = v_{i_j-1} + 1] \left(\frac{1/8 - \Pr[v_{i_j} = v_{i_j-1} - 1]}{\Pr[v_{i_j} = v_{i_j-1} + 1]} \right) \\
&= \Pr[v_{i_j} = v_{i_j-1} - 1] + 1/8 - \Pr[v_{i_j} = v_{i_j-1} - 1] \\
&= 1/8.
\end{aligned}$$

From the standard analysis of the gambler's ruin problem,[2] we have that $(q/p)^{v'_t}$ is a martingale, and thus that the quantity

$$\begin{aligned}
&\Pr[v' \text{ reaches } \sqrt{n} \text{ before } n/8](q/p)^{\sqrt{n}} \\
&\quad + \Pr[v' \text{ reaches } n/8 \text{ before } \sqrt{n}](q/p)^{n/8}
\end{aligned}$$

is equal to $(q/p)^{v_0}$. Because $(q/p)^{n/8} = (1/7)^{n/8}$ is exponentially small, it makes sense to ignore the second addend, leaving

$$\Pr[v' \text{ reaches } \sqrt{n} \text{ before } n/8](q/p)^{\sqrt{n}} < (q/p)^{v_0}$$

or

$$\Pr[v' \text{ reaches } \sqrt{n} \text{ before } n/8] < (q/p)^{v_0 - \sqrt{n}}.$$

[2] See, for example, [8, §XIV.2].

It follows that if $v_0 \geq \sqrt{n} + c\log_7 n$, then the probability that v drops to $\sqrt{n}$ before reaching $n/8$ is bounded by n^{-c}. Once v reaches $n/8$, further drops to $\sqrt{n}$ become exponentially improbable even conditioned on starting at $v = n/8 - 1$. We thus have:

Lemma 8 *Fix $c > 0$. Let $z = o(\sqrt{n})$ and let $v_0 \geq \sqrt{n} + c\log_7 n$. Then for sufficiently large n, the probability that $v_t \leq \sqrt{n}$ for any $t < e^{n/8}n^{-c}$ is less than $2n^{-c}$.*

Proof The probability that v reaches $\sqrt{n}$ before reaching $n/8$ for the first time is at most n^{-c}. For each subsequent drop to $n/8 - 1$, there is a probability of at most $(1/7)^{n/8-1-\sqrt{n}} \leq \exp(-n/8)$ that v reaches $\sqrt{n}$ before returning to $n/8$. Since each such excursion below $n/8$ involves at least one interaction, $e^{n/8}n^{-c}$ interactions give at most an expected n^{-c} drops to $\sqrt{n}$ for a total probability of reaching $v = \sqrt{n}$ bounded by $2n^{-c}$. □

We can apply a similar analysis to the x and y corners, but here the protocol drifts toward the all-x or all-y configuration instead of away from it. Here we track $3y + b$ for the x corner and $3x + b$ for the y corner. Because these functions can change by more than just ± 1, the simple random walk analysis becomes more difficult. Instead, we proceed by showing that $\exp(3y + b)$ is a supermartingale, and bound the probability of moving from $2\sqrt{n}$ to $3\sqrt{n}$ by $\exp(-\sqrt{n})$, the inverse of the change in $\exp(3y + b)$.

Formally, we have:

Lemma 9 *Fix $c > 0$. Let $z = o(\sqrt{n})$ and let $3y_0 + b_0 \leq 2\sqrt{n}$. Then for sufficiently large n, the probability that $3y_t + b_t \geq 3\sqrt{n}$ for any $t < e^{\sqrt{n}-1}n^{-c}$ is less than n^{-c}.*

Proof Let $\sqrt{n} \leq 3y + b \leq 3\sqrt{n}$, so that $x \geq n - O(\sqrt{n})$. We also have $z = o(\sqrt{n}) = o(3y + b) = o(y + b)$.

Examining the change in $3y + b$ in the worst case, we argue as in Lemma 7 to see

$$\Delta(3y + b) = \begin{cases} +2 & \text{w. p. proportional to } (y + z)b, \\ +1 & \text{w. p. proportional to } (y + z)x, \\ -1 & \text{w. p. proportional to } xb, \text{ and} \\ -2 & \text{w. p. proportional to } xy, \end{cases}$$

where "w. p." abbreviates "with probability."

We show that for a suitable constant α independent of n, $\exp(\alpha(3y + b))$ is a supermartingale. We will use the fact that $\exp(t) \leq 1 + t + t^2$ for $|t| \leq 1$.

Let us first consider the expectation of $\exp(\Delta(\alpha(3y + b)))$ conditioned on the event A that the interaction involves an x token as either initiator or responder. We have

$$\begin{aligned} &\mathrm{E}[\exp(\Delta(\alpha(3y + b)))|A] \\ &\quad = \frac{(y + z)e^{\alpha} + be^{-\alpha} + ye^{-2\alpha}}{2y + b + z} \\ &\quad \leq \frac{\begin{pmatrix}(y + z)(1 + \alpha + \alpha^2) \\ + b(1 - \alpha + \alpha^2) \\ + y(1 - 2\alpha + 4\alpha^2)\end{pmatrix}}{2y + b + z} \\ &\quad = \frac{(2y + b + z) + \alpha(z - y - b) + \alpha^2(5y + b + z)}{2y + b + z} \\ &\quad = 1 - \alpha\frac{(1 - o(1))(y + b)}{2y + b + z} + \alpha^2\frac{5y + b + z}{2y + b + z}, \end{aligned}$$

because $z = o(y + b)$. Continuing, this is

$$\begin{aligned} &\leq 1 - \alpha\frac{(1 - o(1))(y + b)}{2y + b + (y/2 + 3b/2)} + \alpha^2\frac{5y + b + z}{2y + b + z} \\ &\leq 1 - \alpha\frac{(1 - o(1))2}{5} + \alpha^2\frac{5y + b + z}{2y + b + z} \\ &\leq 1 - \alpha/3 + \alpha^2\frac{6y + 3b + 3z}{2y + b + z} \\ &= 1 - \alpha/3 + 3\alpha^2. \end{aligned}$$

Setting $\alpha = 1/18$ bounds this quantity by $1 - 1/54 + 3/324 = 107/108 = 1 - 1/108$.

In conditioning on A, we have neglected to include the event B that a y or z token encounters a blank responder. In this latter case, the value of $\exp(\alpha(3y + b))$ rises by a factor of $e^{2\alpha}$. Fortunately, it only occurs with probability proportional to $(y + z)b$, while the event A considered above occurs with probability proportional to $(2y + b + z)x \geq (y + z)x \geq \Omega((y + z)b\sqrt{n})$, since $b = O(\sqrt{n})$. So if we condition on either A or B occurring, we have

$$\begin{aligned} &\mathrm{E}[\exp(\Delta(\alpha(3y + b)))|A \cup B] \\ &\quad = (1 - 1/108)\Pr[A|A \cup B] + \exp(2\alpha)\Pr[B|A \cup B] \\ &\quad = 1 - 1/108 + O(1/\sqrt{n}), \end{aligned}$$

which is strictly less than 1 for sufficiently large n.

If A or B do not occur, $3y + b$ is unchanged; it follows that $\mathrm{E}[\exp(\Delta(\alpha(3y + b)))] \leq 1$ for sufficiently large n, and thus that $\exp((3y_t + b_t)/18)$ is a supermartingale.

Now suppose that at time 0, $\sqrt{n} < 3y + b \leq 2\sqrt{n}$. Let τ be the first time at which $3y + b$ reaches either $\sqrt{n}$ or $3\sqrt{n}$. From the supermartingale property we have that $\mathrm{E}[X_\tau] \leq X_0 \leq \exp(2\alpha\sqrt{n}) = \exp(\sqrt{n}/9)$. Neglecting the $3y + b = \sqrt{n}$ outcome, we have

$$\begin{aligned} \exp(\sqrt{n}/9) &\geq \mathrm{E}[X_\tau] \\ &\geq \Pr[3y_\tau + b_\tau = 3\sqrt{n}]\exp(3\sqrt{n}/18). \end{aligned}$$

from which it follows that

$$\Pr[3y_\tau + b_\tau = 3\sqrt{n}] \leq \exp(-\sqrt{n}/18).$$

We now repeat the excursions argument from the proof of Lemma 8 to obtain the claimed bound. □

7.2 Adjusting the Potential Functions

Our proof of convergence for the non-Byzantine case is based on constructing several potential functions that decrease on average. For the Byzantine case, we keep the same potential functions, but include a new factor that compensates for increases due to Byzantine interactions. We state the general approach in the following lemma:

Lemma 10 *Let f be a function of the states of the non-Byzantine agents and their interaction history in some one-way population protocol, such that f_t is a supermartingale in the absence of Byzantine agents. Let $(f+\Delta f)/f \leq m$ for all transitions involving a Byzantine initiator starting in some subset D of the configuration space. Then $f'_t = f_t m^{-S_t^z}$ is a supermartingale for all times t less than the first time at which the protocol leaves D.*

Proof Let τ be the first time at which the protocol configuration is not in D, and let $t < \tau$. Then either (a) the transition at time t does not include a Byzantine initiator, and we have

$$\begin{aligned}
\mathrm{E}[f'_{t+1}|\mathcal{F}_t] &= \mathrm{E}[f_{t+1}m^{-S^z_{t+1}}|\mathcal{F}_t] \\
&= \mathrm{E}[f_{t+1}m^{-S^z_t}|\mathcal{F}_t] \\
&= \mathrm{E}[f_{t+1}|\mathcal{F}_t]m^{-S^z_t} \\
&\leq f_t m^{-S^z_t} \\
&= f'_t,
\end{aligned}$$

or (b) the transition at time t does include a Byzantine initiator, and we have

$$\begin{aligned}
\mathrm{E}[f'_{t+1}|\mathcal{F}_t] &= \mathrm{E}[f_{t+1}m^{-S^z_{t+1}}|\mathcal{F}_t] \\
&\leq m f_t m^{-S^z_{t+1}} \\
&= m f_t m^{-S^z_t - 1} \\
&= f_t m^{-S^z_t} \\
&= f'_t,
\end{aligned}$$

□

We now produce a common bound $(f+\Delta f)/f \leq 1 + 3/\sqrt{n}$ for *all* of the potential functions used in the proofs of Lemmas 4, 6, and 7, provided we stay within an appropriately-defined domain D.

Lemma 11 *Let one of the following cases hold:*

- *f is $1/(u^2+2n)$ and D is the entire space,*
- *f is $1/v$ and D is all points with $v \geq \sqrt{n}$,*
- *f is $3x+b+1$ and D is all points with $3x+b+1 \geq \sqrt{n}$, or*
- *f is $3y+b+1$ and D is all points with $3y+b+1 \geq \sqrt{n}$.*

Then $(f+\Delta f)/f \leq 1+3/\sqrt{n}$ starting from any point in D.

Proof Observe first that a Byzantine interaction changes the state of at most one normal agent; we use this to find simplified upper bounds on the change in each potential function.

- For $f = 1/(u^2+2n)$, we have

$$\begin{aligned}
\frac{f+\Delta f}{f} &\leq \frac{u^2+2n}{(u-1)^2+2n} \\
&= 1 + \frac{2(u-1)+1}{(u-1)^2+2n} \\
&\leq 1 + 2/\sqrt{n}.
\end{aligned}$$

- For $f = 1/v$, we have

$$\begin{aligned}
\frac{f+\Delta f}{f} &\leq \frac{v}{v-1} \\
&= 1 + \frac{1}{v-1} \\
&\leq 1 + \frac{1}{\sqrt{n}-1} \\
&\leq 1 + 2/\sqrt{n}.
\end{aligned}$$

- For $f = 3x+b+1$, we have

$$\begin{aligned}
\frac{f+\Delta f}{f} &\leq \frac{3(x+1)+b+1}{3x+b+1} \\
&= 1 + \frac{3}{3x+b+1} \\
&\leq 1 + \frac{3}{\sqrt{n}}.
\end{aligned}$$

The case of $f = 3y+b+1$ is symmetric with $f = 3x+b+1$. □

The following lemma bounds the total correction factor:

Lemma 12 *Let $\tau \leq kn\log n$ be a stopping time and let $z = o(n^{1/2})$. Let $m \leq (1+\alpha n^{-1/2})$ for some constant $\alpha > 0$. Then for any fixed $c > 0$,*

$$\Pr[m^{S^z_\tau} \geq n^{o(1)}] \leq n^{-c}.$$

Proof Since $z = o(n^{1/2})$, we have

$$\mathrm{E}[S^z_\tau] = o((n^{1/2}/n)(n\log n)) = o(n^{1/2}\log n),$$

and standard Chernoff bounds show that this bound holds (for a larger constant) without the expectation with probability at least $1 - n^{-c}$. So we have

$$\begin{aligned}
m^{S^z_\tau} &\leq (1+\alpha n^{-1/2})^{o(n^{1/2}\log n)} \\
&= (1+\alpha n^{-1/2})^{(\alpha^{-1}n^{1/2})o(\log n)} \\
&\leq e^{o(\log n)} = n^{o(1)}.
\end{aligned}$$

□

Combining the preceding lemmas and applying the result to our previous bounds for $z = 0$ gives

Corollary 3 *Let k and c be positive constants. Let $\tau \leq \min(\tau_S, kn \log n)$, where τ_S is the first time at which we leave some set of configurations D. Let D be*

- *the entire space for Corollary 2 and Lemma 5,*
- *all points with $v \geq \sqrt{n}$ for Lemma 6,*
- *all points with $3y + b + 1 \geq \sqrt{n}$ for Lemma 7, and*
- *all points with $3x + b + 1 \geq \sqrt{n}$ for the symmetric version of Lemma 7 with y replacing x.*

Then the bounds in each of Corollary 2, Lemma 5, Lemma 6, and Lemma 7 all hold with probability at least $1 - n^{-c+o(1)}$, where all indicator variables are interpreted as taking on the value 0 for Byzantine interactions.

Proof Observe that the proof of each of these bounds is obtained by applying Markov's inequality to the ratio of two potential function values. Any terms in these potential functions involving only indicator variables are unaffected by Byzantine transitions. Subject to the requirement of remaining in the appropriate D for all times prior to τ, it follows from Lemmas 10, 11, and 12 that the remaining quantities $1/(u^2+2n)$, $1/v$, $3x+b+1$, and $3y + b + 1$ are at most $n^{o(1)}$ times their original values, with probability of failure at most n^{-c}, which is easily absorbed into $n^{-c+o(1)}$. This increases the probability of failure obtained from Markov's inequality from n^{-c} to at most $n^{-c+o(1)}$, giving the full result. □

7.3 Convergence Time with Byzantine Agents

Let's put everything together.

Theorem 4 *Let τ be the time at which $x \geq n - \sqrt{n}$, $y \geq n - \sqrt{n}$, or $v \leq \sqrt{n}$ first holds. Let v_0 be the initial number of x's and y's. Then for any fixed $c > 0$ and sufficiently large n, if $v_0 \geq \sqrt{n} + c \log_7 n$, then*

$$\Pr\left[\begin{array}{l}\tau \geq 6769n \log n + 6773cn \log n + 2552n \\ \text{or } v_\tau \leq \sqrt{n}\end{array}\right] = n^{-c+o(1)}. \quad (10)$$

Proof We bound the probabilities of the two events in the union separately. Truncate τ at $10^4 cn \log n$; this does not affect the first bound. Observe that before τ, we lie in the intersection of all four domains D in Corollary 3 (the only tricky part is that we have $3y + b + 1 \geq n - x + 1 \geq \sqrt{n}$ and similarly for $3x + b + 1$). Thus each of the corresponding bounds apply with probability of failure $n^{-c+o(1)}$. Combining these bounds as in the proof of Theorem 1 then gives a probability of τ exceeding the bound in (10) of $5n^{-c+o(1)} = n^{-c+o(1)}$.

For the second event, observe that Lemma 8 applies for times less than $10^4 cn \log n$, so that if τ satisfies the bound we also have that $v_\tau \leq \sqrt{n}$ with probability at most n^{-c}, which is absorbed in $n^{-c+o(1)}$. □

Note that once we are in the x or y corner, Lemma 9 tells us that we remain there with high probability for exponential time. So we have a complete characterization of the convergence behavior of the 3-state majority protocol with $o(\sqrt{n})$ Byzantine agents. It is also not hard to see that the proof of Theorem 2 also continues to hold for $z = o(\sqrt{n})$, since with high probability, the Byzantine agents participate in only $o(\sqrt{n} \log n)$ of the first $O(n \log n)$ interactions, and each interaction involving a Byzantine agent affects the random walk by at most one step.

8 Multi-valued Consensus

It is natural to generalize the problem of reaching consensus on a value of x or y to the problem of reaching consensus on one of m possible input values. There are reductions of multi-valued consensus to binary consensus assuming uniform reliable broadcast [13] or randomization [7]. However, because the criterion we consider is convergence to a single value, in which individual agents may change their decisions and do not in general know when convergence has been achieved, we require a somewhat different reduction.

We describe a bitwise multi-valued consensus protocol that sequentially agrees on the bits of the output value. For m-valued consensus, suppose the input values are represented using $k = \lceil \log_2 m \rceil$ symbols from $\{x, y\}$. The state of each agent is (u, c), where u is a vector of k symbols from $\{x, y\}$ equal to some input value (initially, the input value for this agent) and c is a vector of k symbols from $\{b, x, y\}$ (also initially equal to the input value for this agent.) When an initiator with state (u, c) interacts with a responder with state (u', c'), the protocol computes the new state of the responder as follows.

1. If $c = c'$, then return (u', c') unchanged.
2. Otherwise, let i be the least positive integer such that $c_i \neq c'_i$.
3. If $c_i = b$, then return (u', c') unchanged.
4. If $c'_i \neq b$, then return (u', c'') where $c''_j = c'_j$ for $j = 1, \ldots, i-1$ and $c''_j = b$ for $j = i, \ldots, k$.
5. If $c'_i = b$, then return (u, u).

Lemma 13 *In a population of n agents, with high probability the bitwise consensus protocol converges to a correct consensus value in $O(kn \log n)$ interactions.*

Proof We first observe that throughout the execution of this protocol, if the state of any agent is (u, c), then u is one of the original input values, and c consists of a prefix of u followed by b's. This is true in the initial configuration because for each agent, u and c are set to the input value for this agent. Each interaction preserves this property because steps of type (4) simply increase the length of the suffix of b's in c, and steps of type (5) set both components to u, which is one of the original input values in the computation.

Clearly, stable configurations are those in which all agents have the same value of c. Consider any reachable configuration in which all the agents agree on the values of c_j for $j = 1, \ldots, i-1$. Then no interaction will change these values, and considering the values of c in position i, the protocol is following the steps of the approximate majority protocol exactly. Thus, with high probability, convergence to agreement on the value of c_i will occur within $O(n \log n)$ interactions. Hence, by induction on $i = 1, \ldots, k$, with high probability, convergence to agreement on the value of c will occur withing $O(kn \log n)$ interactions.

To see that the value of c agreed upon is in fact one of the original input values, we note that if (u, c) is the state of some agent, then u must be one of the original input values and c must be a prefix of u followed by b's. To see that c will in fact contain no b's we argue as follows. Suppose there is a reachable configuration in which all the values of c are the same and c has a non-empty suffix of $s > 0$ b's. Consider a shortest execution in which this configuration is reached, and consider the interaction that reaches it. The initiator must already have c, and the responder must change to c from some other value. That previous value cannot have a blank suffix shorter than or equal to that of c; the only possibility is a shorter non-blank prefix of c followed by blanks. In this case, a step of type (5) occurs, and the resulting value of c has no blanks at all, which is a contradiction. Thus, the protocol cannot converge to a value of c with any blanks, and the final value of c must be one of the original input values. □

It is open whether this protocol also has useful majority or robustness properties.

9 Acknowledgments

The second author would like to thank Joanna Ellman-Aspnes for a suggestion that helped overcome an obstacle in the proof of Lemma 4. The authors would like to thank the DISC 2007 reviewers and the referees of the present paper for their helpful comments.

Part of this work was done while the third author was a student at Princeton University.

James Aspnes was supported in part by NSF grant CNS-0435201. David Eisenstat was supported in part by a Gordon Y. S. Wu Fellowship and a National Defense Science and Engineering Graduate Fellowship.

References

1. Alon, N., Spencer, J.H.: The Probabilistic Method. John Wiley & Sons (1992)
2. Angluin, D., Aspnes, J., Diamadi, Z., Fischer, M.J., Peralta, R.: Computation in networks of passively mobile finite-state sensors. Distributed Computing pp. 235–253 (2006)
3. Angluin, D., Aspnes, J., Eisenstat, D.: Fast computation by population protocols with a leader. In: Distributed Computing: 20th International Symposium, DISC 2006: Stockholm, Sweden, September 2006: Proceedings, pp. 61–75 (2006)
4. Angluin, D., Aspnes, J., Eisenstat, D., Ruppert, E.: The computational power of population protocols. Distributed Computing **20**(4), 279–304 (2007)
5. Aspnes, J., Ruppert, E.: An introduction to population protocols. Bulletin of the European Association for Theoretical Computer Science **93**, 98–117 (2007)
6. Chow, Y.S., Robbins, H., Siegmund, D.: The Theory of Optimal Stopping. Dover (1991)
7. Ezhilchelvan, P., Mostefaoui, A., Raynal, M.: Randomized multivalued consensus. In: ISORC '01: Proceedings of the Fourth International Symposium on Object-Oriented Real-Time Distributed Computing, p. 195. IEEE Computer Society, Washington, DC, USA (2001)
8. Feller, W.: An Introduction to Probability and its Applications, vol. 1, third edn. John Wiley and Sons (1958)
9. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry **81**(25), 2340–2361 (1977)
10. Gillespie, D.T.: A rigorous derivation of the chemical master equation. Physica A **188**, 404–425 (1992)
11. Grimmet, G.R., Stirzaker, D.R.: Probability and Random Processes, 2nd edn. Oxford Science Publications (1992)
12. Kurtz, T.G.: Approximation of Population Processes. No. 36 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia (1981)
13. Mostefaoui, A., Raynal, M., Tronel, F.: From binary consensus to multivalued consensus in asynchronous message-passing systems. Information Processing Letters **73**(5-6), 2007–212 (2000)
14. Wormald, N.C.: Differential equations for random processes and random graphs. Annals of Applied Probability **5**(4), 1217–1235 (1995)