The Availability of Crumbling Wall Quorum Systems*

David Peleg† Avishai Wool‡

December 25, 1995

Abstract

A *quorum system* is a collection of sets (quorums) every two of which intersect. Quorum systems have been used for many applications in the area of distributed systems, including mutual exclusion, data replication and dissemination of information.

Crumbling walls are a general class of quorum systems. The elements (processors) of a wall are logically arranged in *rows* of varying *widths*. A quorum in a wall is the union of one full row and a representative from every row below the full row. This class considerably generalizes a number of known quorum system constructions.

In this paper we study the availability of crumbling wall quorum systems. We show that if the row width is bounded, or if the number of rows is bounded, then the wall's failure probability F_p does not vanish as the number of elements tends to infinity (i.e., F_p is not Condorcet). If the wall may grow in both the row number and row width, we show that the behavior depends on the *rate* of growth of the row width. We establish a sharp threshold rate: when the row width $n_i \leq \lfloor \log_2 2i \rfloor$ then F_p *is* Condorcet, and when $n_i \geq (1+\varepsilon)\log_2 i$ then F_p *is not* Condorcet.

*An extended abstract of this work has appeared in the 14'th ACM Symp. Princip. of Dist. Comp., 1995.

†Department of Applied Mathematics and Computer Science, The Weizmann Institute, Rehovot 76100, Israel. Supported in part by a Walter and Elise Haas Career Development Award and by a grant from the Israeli Basic Research Foundation. E-mail: peleg@wisdom.weizmann.ac.il.

‡Department of Applied Mathematics and Computer Science, The Weizmann Institute, Rehovot 76100, Israel. E-mail: yash@wisdom.weizmann.ac.il.

1 Introduction

1.1 Motivation

Quorum systems serve as a basic tool providing a uniform and reliable way to achieve coordination between processors in a distributed system. Quorum systems are defined as follows. A *set system* is a collection of sets $\mathcal{S} = \{S_1, \ldots, S_m\}$ over an underlying universe $U = \{u_1, \ldots, u_n\}$. A set system is said to satisfy the *intersection property*, if every two sets $S, R \in \mathcal{S}$ have a nonempty intersection. Set systems with the intersection property are known as *quorum systems*, and the sets in such a system are called quorums.

Quorum systems have been used in the study of distributed control and management problems such as *mutual exclusion* (cf. [Ray86]), *data replication protocols* (cf. [DGS85, Her84]), *name servers* (cf. [MV88]), *selective dissemination of information* (cf. [YG94]), and *distributed access control and signatures* (cf. [NW96]).

A protocol template based on quorum systems works as follows. In order to perform some action (e.g., update the database, enter a critical section), the user selects a quorum and *accesses all its elements*. The intersection property then guarantees that the user will have a consistent view of the current state of the system. For example, if all the members of a certain quorum give the user permission to enter the critical section, then any other user trying to enter the critical section before the first user has exited (and released the permission-granting quorum from its lock) will be refused permission by at least one member of any quorum it chooses to access.

A well studied measure of the quality of a quorum system is its *Availability*. Assuming that each element fails with probability p, what is the probability, F_p, that the surviving elements do not contain any quorum? This failure probability measures how resilient the system is, and we would like F_p to be as small as possible. A desirable asymptotic behavior of F_p is that $F_p \to 0$ when $n \to \infty$ for *all* $p < \frac{1}{2}$, and such an F_p is called Condorcet, after [Con].

The *Crumbling Walls* class of quorum system constructions was introduced in [PW95b]. The construction is defined as follows. The elements are arranged in *rows*, and a quorum is the union of one full row and a single representative from every row below the full row. No restriction is placed on the row widths, and the "wall" is allowed to crumble at its edge (see Figure 1). Formally,

Definition 1.1 (Crumbling Wall) *Let* $\boldsymbol{n} = (n_1, \ldots, n_d)$ *be such that* $\sum_{i=1}^{d} n_i = n$. *Let* $U_1, \ldots, U_d$ *be nonempty disjoint subsets of the universe* U *with* $|U_i| = n_i$. *Then*

$$\mathrm{CW}\langle \boldsymbol{n} \rangle = \Big\{ U_i \cup \{u_{i+1}, \ldots, u_d\} : u_j \in U_j \ \text{ for } \ j = i+1, \ldots, d \Big\}$$

is the crumbling wall defined by $\boldsymbol{n}$. *The set* U_i *is called the* i*'th* row *and* n_i *is its* width.

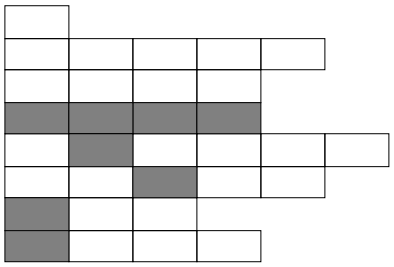

Figure 1: The crumbling wall $\mathrm{CW}\langle 1,5,4,4,6,5,3,4\rangle$, with one quorum shaded.

Crumbling walls generalize a number of known quorum system constructions, including: (i) The singleton system Sngl, which is a trivial wall with $\boldsymbol{n} = (1)$. (ii) The Wheel [MP92, PW95a], which is the wall defined by $\boldsymbol{n} = (1, n-1)$. (iii) The Grid of [CAA90], which is a wall defined by $\boldsymbol{n} = (d, d, \ldots, d)$,[1] and the hollow grids of [KRS93] which can be represented similarly. (iv) The triangular system [Lov73, EL75], denoted by Triang, which is a wall defined by $\boldsymbol{n} = (1, 2, \ldots, d)$. (v) The Lovász coteries of [Nei92], which are walls with $n_1 = 1$ and $n_i \geq 2$ for all $i \geq 2$.

Special emphasis is given in [PW95b] to the CWlog system, which is a wall with row width $n_i = \lfloor \lg 2i \rfloor$.[2] It is shown that CWlog has many advantageous properties, such as log-sized quorums and low load. Calculations demonstrate that CWlog has high availability for small systems, with $n \leq 100$.

In this paper we study the availability of crumbling walls systems. Our emphasis is on the asymptotic behavior of the failure probability F_p, and in particular we investigate when F_p is or is not Condorcet. The asymptotic availability of the Sngl, Triang, Grid and Wheel crumbling walls has been analyzed in [KC91, RST92, PW95a]. All four of these constructions share the property that their failure probabilities are *not* Condorcet, i.e., when n increases, F_p does not vanish when $p < \frac{1}{2}$. Therefore it is somewhat surprising that this is not true for all wall families. Specifically, we prove that the CWlog wall has a failure probability that *is* Condorcet.

1.2 Related Work

The first distributed control protocols using quorum systems [Tho79, Gif79] use *voting* to define the quorums. Each processor has a number of votes, and a quorum is any set of processors with a combined number of votes exceeding half of the system's total number of votes. The simple majority system is the most obvious voting system.

[1] Usually a quorum in a Grid is one full row and a representative in *every* other row. Our somewhat improved variant, in which representatives are required only *below* the full row, has smaller quorums and dominates the regular Grid.

[2] We use lg to denote $\log_2$ and ln to denote $\log_e$.

The availability of voting systems is studied in [BG87]. It is shown that in terms of availability, the majority is the best quorum system when $p < \frac{1}{2}$. In [PW95a, DKK+94] the failure probability function F_p is characterized, and among other things it is shown that the singleton has the best availability when $p > \frac{1}{2}$. The case when the elements fail with different probabilities p_i, all less than $\frac{1}{2}$, is addressed in [SB94].

The first paper to explicitly consider mutual exclusion protocols in the context of intersecting set systems is [GB85]. In this work the term *coterie* and the concept of *domination* are introduced. Several basic properties of dominated and non-dominated coteries are proved.

Alternative protocols based on quorum systems (rather than on voting) appear in [Mae85] (using finite projective planes), [AE91] (the Tree system), [CAA90, KRS93] (using a grid), [Kum91, KC91, RT91, RST92] (hierarchical systems). The triangular system is due to [Lov73, EL75]. A generalization of the triangular system appears in [Nei92] under the name Lovász coteries. The Wheel system appears in [MP92].

In [HMP95], the question of how evenly balanced the work load can be is studied. Tradeoffs between the potential load balancing of a system and its average load are obtained. The notion of load is studied further in [NW94]. Lower bounds on the load and tradeoffs between the load and availability are shown. Four quorum system constructions are shown, featuring optimal load and high availability.

1.3 Contents

In this paper we analyze the asymptotic behavior of the failure probability of crumbling walls. Since the wall system over a universe of size n is not unique, we have some freedom in choosing the way the construction scales up as n increases. We show that if the row width is bounded, or if the number of rows is bounded, then F_p is not Condorcet. If the wall may grow in both the row number and row width, we show that the behavior depends on the *rate* of growth of the row width. We establish a sharp threshold rate: when the row width $n_i \leq \lfloor \lg 2i \rfloor$ then F_p *is* Condorcet, and when $n_i \geq (1+\varepsilon)\lg i$ then F_p *is not* Condorcet. An important part of the analysis is a proof that the CWlog system has a Condorcet failure probability. Moreover, our results show that the CWlog system is essentially the *only* high availability crumbling wall.

The organization of this paper is as follows. In Section 2 we introduce the definitions and notation, and list some useful theorems. In Section 3 we specify the assumptions we make on the structure of the walls, and prove some immediate consequences. In Section 4 we deal with bounded walls. In Section 5 we consider "thick" walls, with row width $n_i \geq (1+\varepsilon)\lg i$. In Section 6 we consider "thin" walls, that have row widths of $n_i \leq \lfloor \lg 2i \rfloor$ (with some minor restrictions to be defined later on). In particular we prove that the CWlog system has Condorcet

failure probability.

2 Preliminaries

2.1 Definitions and Notation

Let us first define the basic terminology used later on.

Definition 2.1 *A* Set System $\mathcal{S} = \{S_1, \ldots, S_m\}$ *is a collection of subsets* $S_i \subseteq U$ *of a finite universe* U. *A* Quorum System *is a set system* $\mathcal{S}$ *that has the* Intersection property: $S \cap R \neq \varnothing$ *for all* $S, R \in \mathcal{S}$.

Alternatively, quorum systems are known as *intersecting set systems* or as *intersecting hypergraphs*. The sets of the system are called *quorums*. The number of elements in the underlying universe is denoted by $n = |U|$.

Definition 2.2 *A* Coterie *is a quorum system* $\mathcal{S}$ *that has the* Minimality property: *there are no* $S, R \in \mathcal{S}$, *s.t.* $S \subset R$.

Definition 2.3 *Let* $\mathcal{R}, \mathcal{S}$ *be coteries (over the same universe* U*). Then* $\mathcal{R}$ dominates $\mathcal{S}$, *denoted* $\mathcal{R} \succ \mathcal{S}$, *if* $\mathcal{R} \neq \mathcal{S}$ *and for each* $S \in \mathcal{S}$ *there is* $R \in \mathcal{R}$ *such that* $R \subseteq S$. *A coterie* $\mathcal{S}$ *is called* dominated *if there exists a coterie* $\mathcal{R}$ *such that* $\mathcal{R} \succ \mathcal{S}$. *If no such coterie exists then* $\mathcal{S}$ *is* non-dominated (ND). *Let* NDC *denote the class of all* ND *coteries.*

The following proposition of [Nei92] and [PW95b] shows that the Lovász coteries of [Nei92] are the *only* ND walls constructions.

Proposition 2.4 $\mathrm{CW}\langle \boldsymbol{n} \rangle \in \mathrm{NDC}$ *iff* $n_1 = 1$ *and* $n_i \geq 2$ *for all* $2 \leq i \leq d$.

2.2 The Probabilistic Failure Model

We use a simple probabilistic model of the failures in the system. We assume that the elements (processors) fail independently with a fixed uniform probability p. We assume that the failures are *transient*, that the failures are *crash* failures (i.e., a failed element stops to function rather than functions incorrectly), and that they are *detectable*.

Notation: We use $q = 1 - p$ to denote the probability of an element survival.

Definition 2.5 *For every quorum* $S \in \mathcal{S}$ *let* $\mathcal{E}_S$ *be the event that* S *is* hit, *i.e., at least one element* $i \in S$ *has failed. Let* $\mathit{fail}(\mathcal{S})$ *be the event that all the quorums* $S \in \mathcal{S}$ *were hit, i.e.,* $\mathit{fail}(\mathcal{S}) = \bigcap_{S \in \mathcal{S}} \mathcal{E}_S$.

We can now define the global system failure probability of a quorum system $\mathcal{S}$ as follows.

Definition 2.6 $F_p(\mathcal{S}) = \mathbb{P}(fail(\mathcal{S})) = \mathbb{P}\left(\bigcap_{S\in\mathcal{S}} \mathcal{E}_S\right)$.

The following theorems of [PW95a] describe some properties of the failure probability F_p.

Theorem 2.7 *A coterie $\mathcal{S}$ is* ND *iff* $F_{1/2}(\mathcal{S}) = \frac{1}{2}$.

Theorem 2.8 (Symmetry) *For any* $\mathcal{S} \in$ NDC, $F_p(\mathcal{S}) + F_{1-p}(\mathcal{S}) = 1$.

Proposition 2.9 *$F_p(\mathcal{S})$ is strictly increasing with p for every quorum system $\mathcal{S}$.*

When we consider the asymptotic behavior of $F_p(\mathcal{S}_n)$ for a sequence $\mathcal{S}_n$ of quorum systems over a universe with an increasing size n, we find that for many constructions it is similar to the behavior described by the Condorcet Jury Theorem [Con]. Hence, the following definition of [PW95a].

Definition 2.10 *A parameterized family of functions $g_p(n) : \mathbb{N} \to [0,1]$, for $p \in [0,1]$, is said to be* Condorcet *iff* $\lim_{n\to\infty} g_p(n) = \begin{cases} 0, & p < \frac{1}{2}, \\ 1, & p > \frac{1}{2}, \end{cases}$ *and $g_{1/2}(n) = \frac{1}{2}$ for all n.*

Below we list some of the basic results of [PW95b] which we need in the asymptotic analysis. These results give formulas for the failure probability of a crumbling wall, and show that walls with monotone increasing row widths have the best availability among all the row permutations.

Notation: Let $F_p(i)$ denote F_p of the sub-wall of the top i rows of $\mathrm{CW}\langle \boldsymbol{n} \rangle$.

Fact 2.11 [PW95b] *Let $\mathrm{CW}\langle \boldsymbol{n} \rangle$ be given. Then the sub-wall failure probability $F_p(i)$ obeys the recurrence*

$$\begin{cases} F_p(1) = 1 - q^{n_1}, \\ F_p(i) = p^{n_i} + (1 - p^{n_i} - q^{n_i}) F_p(i-1), \quad i > 1. \end{cases}$$

Fact 2.12 [PW95b] *The failure probability of a wall $\mathrm{CW}\langle \boldsymbol{n} \rangle$ on d rows with $n_1 = 1$ is*

$$F_p(\mathrm{CW}\langle \boldsymbol{n} \rangle) = \sum_{i=1}^{d} p^{n_i} \prod_{j=i+1}^{d} \left(1 - p^{n_j} - q^{n_j}\right).$$

Proposition 2.13 [PW95b] *Out of all the walls defined by some permutation of $(n_1, \ldots, n_d)$, the wall with the minimal failure probability when $0 < p < \frac{1}{2}$ has its rows in a monotone non-decreasing order of widths.* ∎

3 Basic Properties

3.1 Assumptions

Since the wall system over a universe of size n is not unique, analyzing its asymptotic behavior requires us to restrict ourselves to some specific subclasses of walls.

Denote an infinite sequence of walls by $(W_1, W_2, \ldots)$ where $W_t = \mathrm{CW}\langle n_1^t, n_2^t, \ldots, n_{d_t}^t \rangle$. In light of Propositions 2.4 and 2.13, we require the sequence (W_t) to obey the following assumptions.

Assumption 3.1 *All the walls W_t in the sequence are Lovász coteries defined by non-decreasing width sequences, i.e., $1 = n_1^t < n_2^t \leq \cdots \leq n_{d_t}^t$.*

Assumption 3.2 *Rows do not "shrink" in width, i.e., once row i reaches a width n_i^t in W_t, the t'th wall of the sequence, then $n_i^{t'} \geq n_i^t$ for all $t' > t$. Note that this implies that the universe size $n = \sum_{i=1}^{d_t} n_i^t$ increases with t.*

In most of the analysis we also require the sequence to obey the following restriction.

Assumption 3.3 *Rows neither "shrink" nor "expand", i.e., once row i exists in W_t, then $n_i^{t'} = n_i^t$ for all $t' > t$.*

Note that whenever Assumption 3.3 holds, we may drop the superscript and speak of n_i, the width of row i, as it does not change with t.

Unless otherwise noted, we require the sequence (W_t) to obey Assumptions 3.1–3.3. In particular, we assume that the i'th row has the same width n_i in all the walls W_t that have at least i rows. Therefore, when every wall W_t has $d_t = t$ rows then the sequence of walls (W_t) is fully characterized by the sequence of numbers (n_i). This is formalized through the following notation, used to describe the class of infinite families of walls we are interested in.

Notation: An infinite sequence of integers (n_i) is called a *standard* sequence if it is non-decreasing, with $n_1 = 1$ and $n_i \geq 2$ for all $i > 1$. Let $\boldsymbol{n}^{[d]}$ denote the prefix $n_1, \ldots, n_d$ of the sequence, and let $\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle$ be the wall with row widths $n_1 \ldots n_d$.

The sequences we will be looking at throughout most of the paper are of the form $(W_d) = \mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle$.

3.2 Adding Rows Improves the Availability

The following lemma sets a bound on F_p of a wall with $d-1$ rows in terms of the width of the *next* row, i.e., the row that appears in $\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle$ but not in $\mathrm{CW}\langle \boldsymbol{n}^{[d-1]} \rangle$. This lemma allows us to prove (in Proposition 3.5) that adding rows at the bottom improves the availability, and is also useful later on.

Lemma 3.4 *Let (n_i) be a standard sequence. If $0 < p < \frac{1}{2}$ and $d \geq 2$ then*

$$F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d-1]} \rangle) > \frac{p^{n_d}}{p^{n_d} + q^{n_d}} \ .$$

Proof: As in Fact 2.11, let $F_p(d)$ denote $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$. We prove the claim by induction on d. For the induction base ($d = 2$) we need to show that $F_p(1) = p > \frac{p^{n_2}}{p^{n_2}+q^{n_2}}$. Note that the function $p^x/(p^x + q^x)$ is strictly decreasing with x when $p < \frac{1}{2} < q$, so we can assume that $n_2 = 2$. A simple check shows that indeed $p > p^2/(p^2 + q^2)$ when $0 < p < \frac{1}{2}$.

For the inductive step, assume the claim holds up to $d-1$. Then using Fact 2.11,

$$\begin{aligned} F_p(d) &= p^{n_d} + (1 - p^{n_d} - q^{n_d})F_p(d-1) \\ &> p^{n_d} + (1 - p^{n_d} - q^{n_d}) \cdot \frac{p^{n_d}}{p^{n_d} + q^{n_d}} = \frac{p^{n_d}}{p^{n_d} + q^{n_d}} \geq \frac{p^{n_{d+1}}}{p^{n_{d+1}} + q^{n_{d+1}}} \ . \end{aligned}$$

The last step holds since $n_{d+1} \geq n_d$. ∎

Remark: We can bound $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d-1]} \rangle)$ in terms of n_{d-1} instead of n_d by an almost identical proof. However the inequality is not strict in this case, and therefore is less useful.

The next proposition shows that adding rows at the bottom improves the availability of the wall.

Proposition 3.5 *Let (n_i) be a standard sequence. If $0 < p < \frac{1}{2}$ then $F_p(d) = F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ is a strictly decreasing sequence.*

Proof: We need to show that $F_p(d-1) > F_p(d)$ for $d \geq 2$. Using Fact 2.11 this amounts to showing that

$$F_p(d-1) > p^{n_d} + (1 - p^{n_d} - q^{n_d})F_p(d-1) \ ,$$

which is a direct consequence of Lemma 3.4. ∎

4 Bounded Walls

4.1 Walls with a Bounded Number of Rows

In this section we deal with wall families with a bounded number of rows. This is the only part of this paper where the wall sequence obeys Assumptions 3.1–3.2 but not Assumption 3.3, and rows' widths do in fact increase. The Wheel system [MP92, PW95a] is an example of such a family, in which there are precisely 2 rows in every wall, and as the universe size increases so does the width of the second row. The following proposition characterizes the asymptotics of F_p in this case.

Proposition 4.1 *Let $(W_1, W_2, \ldots)$, where $W_t = \mathrm{CW}\langle n_1^t, n_2^t, \ldots, n_{d_t}^t \rangle$, be an infinite family of walls over universes of increasing size, such that the number of rows is bounded by a constant d. Then for any $0 < p < 1$ there exists $c_p > 0$ such that $F_p(W_t) \geq c_p$ for all sufficiently large t. In other words, $F_p(W_t)$ is not Condorcet.*

Proof: Since we are concerned with asymptotics, we can truncate the sequence (W_t) by dropping the prefix consisting of all the walls with less than d rows. Therefore we can assume that all the walls have exactly d rows. By Assumptions 3.1–3.2, there exists a number $1 \leq \ell \leq d-1$ such that the first ℓ rows have bounded widths, and all the rows $i > \ell$ have widths n_i^t tending to infinity with t ($\ell \geq 1$ since $n_1^t = 1$ for all t). Again we can assume that rows $1, \ldots, \ell$ have reached their final width (by truncating the sequence and dropping the prefix in which these widths have not yet been attained), and that the walls increase in width only in rows $\ell+1, \ldots, d$. Let $\mathcal{B} = \mathrm{CW}\langle n_1, \ldots, n_\ell \rangle$ denote the (fixed) wall of the first ℓ rows. Then by Fact 2.11, for any $0 < p < 1$,

$$\begin{aligned} F_p(W_t) &= F_p(\mathcal{B}) \prod_{j=\ell+1}^{d} (1 - p^{n_j^t} - q^{n_j^t}) + \sum_{i=\ell+1}^{d} \left(p^{n_i^t} \prod_{j=i+1}^{d} (1 - p^{n_j^t} - q^{n_j^t}) \right) \\ &\underset{t \to \infty}{\longrightarrow} F_p(\mathcal{B}) > 0, \end{aligned}$$

since d is a constant, so the second term vanishes and the product in the first term tends to 1. The claim follows. ∎

4.2 Walls with a Bounded Row Width

Next we analyze the case of a family of walls with an unbounded number of rows, but with a bounded row *width*.

Proposition 4.2 *Let (n_i) be a bounded standard sequence, and let $k = \max_i\{n_i\}$. Then $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle) \underset{d \to \infty}{\longrightarrow} \frac{p^k}{p^k + q^k}$ for any $0 < p < 1$, so $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ is not Condorcet.*

Proof: By the premise the width k is reached, i.e., there exists a finite i_0 such that $n_i = k$ for all $i > i_0$. To see immediately that $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ does not vanish, note that if $d \geq i_0$ then $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle) > p^k/(p^k + q^k) > 0$ by Lemma 3.4. In fact, the bound is achieved asymptotically, since by Fact 2.12

$$\begin{aligned} &F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle) = \\ &\sum_{i=1}^{i_0} p^{n_i} \prod_{j=i+1}^{i_0} (1 - p^{n_j} - q^{n_j}) \left[(1 - p^k - q^k)^{d-i_0} \right] + p^k \sum_{i=i_0+1}^{d} (1 - p^k - q^k)^{d-i} \\ &\quad \underset{d \to \infty}{\longrightarrow} \frac{p^k}{p^k + q^k}, \end{aligned}$$

since the first sum vanishes and the second is a geometric series. ∎

5 Walls with Super-Logarithmic Row Widths

In this section we consider walls with row widths that increase at a super-logarithmic rate, namely $n_i \geq (1+\varepsilon)\lg i$. The main claim in this section (Theorem 5.3) is that walls with super-logarithmic row widths do not have Condorcet failure probabilities. The Triang system [Lov73, EL75] is an example of a wall family with super-logarithmic row widths.

We start by looking at the behavior of F_p for a fixed value of p.

Lemma 5.1 *Let (n_i) be a standard sequence, and let $0 < p < 1$ and $\varepsilon > 0$ be given. If $n_i \geq (1+\varepsilon)\log_{1/q} i$ for all i then $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle) \geq \varepsilon'(p)$ for some $\varepsilon'(p) > 0$.*

Proof: If all the rows are hit (namely, contain a failed element), then the system certainly fails. Therefore,

$$F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle) \geq \mathbb{P}(\text{all rows are hit}) = \prod_{i=1}^{d} \left(1 - q^{n_i}\right).$$

Note that the function e^{-cx}, for $c > 1$, crosses the function $1-x$ twice, at $x = 0$ and at $x = \alpha$ for some $0 < \alpha < 1$ that satisfies $c = \frac{1}{\alpha}\ln\frac{1}{1-\alpha}$. Moreover, $1 - x \geq e^{-cx}$ in the range $0 \leq x \leq \alpha$. Setting $\alpha = q$ (the success probability), if $c = \frac{1}{q}\ln\frac{1}{p}$ then $1 - q^{n_i} \geq e^{-cq^{n_i}}$ for all $i \geq 1$ since $0 < q^{n_i} < q$. Therefore

$$\prod_{i=1}^{d} \left(1 - q^{n_i}\right) \geq e^{-c\sum_{i=1}^{d} q^{n_i}} > e^{-c\sum_{i=1}^{\infty} q^{n_i}},$$

and if $n_i \geq (1+\varepsilon)\log_{1/q} i$ then $q^{n_i} \leq \left(\frac{1}{i}\right)^{1+\varepsilon}$ so the series converges and we are done. ∎

Recall that by Proposition 2.9, F_p is an increasing function of p, so if F_p does not vanish at $p = \beta$ then it does not vanish for any $p \geq \beta$ either. Furthermore, by Theorem 2.7, F_p does not vanish for $p \geq \frac{1}{2}$ for any quorum system. Note also that Lemma 5.1 gives a different row width sequence for every p. Therefore, in the following corollary we rephrase Lemma 5.1 by looking at a *given* sequence and giving conditions for F_p not to vanish in the interesting range $0 < p < \frac{1}{2}$.

Corollary 5.2 *Let $\varepsilon > 0$ be given. If there exists $0 < \beta < \frac{1}{2}$ such that $n_i \geq (1+\varepsilon)\log_{\frac{1}{1-\beta}} i$ for all i, then $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ does not vanish for $\beta \leq p \leq 1$.*

The following theorem shows that "thick" walls, with $n_i \geq (1+\varepsilon)\lg i$ do not have a Condorcet failure probability.

Theorem 5.3 *Let $\varepsilon' > 0$ be given. If $n_i \geq (1+\varepsilon')\lg i$ for all i, then there exists $\delta > 0$ such that $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ does not vanish for $p \in [\frac{1}{2} - \delta, \frac{1}{2})$, so $F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle)$ is not Condorcet.*

Proof: There exists $\delta > 0$ small enough that $(1+\varepsilon')\lg\left(\frac{2}{1+2\delta}\right) > 1$. For this δ, set $\varepsilon =$

$(1+\varepsilon')\lg\left(\frac{2}{1+2\delta}\right)-1$. Then for all i,

$$n_i \geq (1+\varepsilon')\lg i = \frac{1+\varepsilon}{\lg\left(\frac{2}{1+2\delta}\right)}\lg i = (1+\varepsilon)\log_{\frac{2}{1+2\delta}} i.$$

Setting $\beta = \frac{1}{2}-\delta$ and applying Corollary 5.2 completes the proof. ∎

Remark: A result stronger than Theorem 5.3 holds for "very thick" walls. In [PW95a] it is shown that $F_p(\text{Triang}) \geq p^{1/p}$ for any $0 \leq p \leq 1$, so $F_p(\text{Triang})$ does not vanish for *any* $p \neq 0$, rather than only near $p = \frac{1}{2}$. The proof works for any wall that is as thick as the Triang (i.e., $n_i \geq i$).

6 Walls with Sub-Logarithmic Row Widths

We now prove that in contrast to all the other cases we have seen so far, F_p of "thin" wall families *is* Condorcet. We start by showing in Theorem 6.1 that the CWlog system has a Condorcet failure probability, with $F_p(\text{CWlog}) = O((\frac{\lg n}{n})^{\varepsilon})$ for some $\varepsilon = \varepsilon(p) > 0$ and for all $0 < p < \frac{1}{2}$. This serves us in several ways. First it shows that as we claimed before, CWlog has high asymptotic availability. Secondly, it shows that the logarithmic criterion in Theorem 5.3 is tight. Finally, our claim that F_p of wall families with $n_i \leq \lfloor \lg 2i \rfloor$ is Condorcet (Theorem 6.4) is proved by comparing some arbitrary wall to one that is known to have a Condorcet failure probability. For this we need the example of Theorem 6.1.

Theorem 6.1 *Consider the* CWlog *system on d rows, with $n_i = \lfloor \lg 2i \rfloor$, and let α be such that $\alpha + \lg(1/\alpha) = 2$ ($\alpha \approx 0.3099$). Then*

$$F_p(\text{CWlog}) \leq \begin{cases} C_1\left(\frac{1}{d}\right)^q, & 0 < p < \alpha, \\ C_2\frac{\log d}{d^q}, & p = \alpha, \\ C_3\left(\frac{1}{d}\right)^{(\lg\frac{1}{p}-1)}, & \alpha < p < \frac{1}{2}, \end{cases}$$

for some C_1, C_2, C_3 that depend only on p. Therefore $F_p(\text{CWlog}) \underset{d\to\infty}{\longrightarrow} 0$ for all $0 < p < \frac{1}{2}$, thus $F_p(\text{CWlog})$ is Condorcet.

Proof: By Fact 2.12, $F_p(d) = F_p(\text{CW}\langle \boldsymbol{n}^{[d]} \rangle) = \sum_{i=1}^{d} p^{n_i} \prod_{j=i+1}^{d} \left(1-p^{n_j}-q^{n_j}\right)$. We first estimate the product. Since $1 - x \leq e^{-x}$,

$$\prod_{j=i+1}^{d}\left(1-p^{n_j}-q^{n_j}\right) \leq \prod_{j=i+1}^{d}\left(1-q^{n_j}\right) \leq e^{-\sum_{j=i+1}^{d} q^{n_j}}. \qquad (1)$$

By the definition, $n_j \leq \lg j + 1$, therefore $q^{n_j} \geq q(\frac{1}{j})^{\lg(1/q)}$. Note that $\lg(1/q) < 1$ when $q > \frac{1}{2}$. Therefore

$$\begin{aligned}\sum_{j=i+1}^{d} q^{n_j} &\geq q \sum_{j=i+1}^{d} \left(\frac{1}{j}\right)^{\lg(1/q)} \geq q \sum_{j=i+1}^{d} \left(\frac{1}{j}\right) \\ &\geq q \int_{i}^{d} \frac{dx}{x+1} = q\left(\ln(d+1) - \ln(i+1)\right),\end{aligned}$$

and

$$e^{-\sum_{j=i+1}^{d} q^{n_j}} \leq e^{-q(\ln(d+1)-\ln(i+1))} = \left(\frac{i+1}{d+1}\right)^{q}. \tag{2}$$

Using (1) and (2) we obtain that

$$F_p(d) \leq \sum_{i=1}^{d} p^{n_i} \left(\frac{i+1}{d+1}\right)^{q} = \left(\frac{1}{d+1}\right)^{q} \sum_{i=1}^{d} (i+1)^q p^{n_i}. \tag{3}$$

Since $n_i \geq \lg i$, $p^{n_i} \leq (1/i)^{\lg(1/p)}$. Note that $\lg(1/p) > 1$ when $p < \frac{1}{2}$. Plugging this into (3) we get

$$F_p(d) \leq \left(\frac{1}{d+1}\right)^{q} \sum_{i=1}^{d} (2i)^q \left(\frac{1}{i}\right)^{\lg(1/p)} = \left(\frac{2}{d+1}\right)^{q} \sum_{i=1}^{d} \left(\frac{1}{i}\right)^{\lg(1/p)-q}. \tag{4}$$

Denote $\gamma = \lg(1/p) - q$. Then $\gamma = 1$ when $p = \alpha \approx 0.3099$. We observe three cases:

- $\gamma > 1$, thus $\sum_{i=1}^{d} \left(\frac{1}{i}\right)^{\gamma} < C_1$ and $F_p(d) < C_1 \left(\frac{2}{d+1}\right)^{q}$,
- $\gamma = 1$, thus $\sum_{i=1}^{d} \left(\frac{1}{i}\right)^{\gamma} = O(\log d)$ and $F_p(d) < C_2 \frac{\log d}{d^q}$,
- $1 - q < \gamma < 1$, thus $\sum_{i=1}^{d} \left(\frac{1}{i}\right)^{\gamma} \leq 1 + \int_{1}^{d} \frac{dx}{x^{\gamma}} = O(d^{1-\gamma})$ and $F_p(d) < C_3 \frac{d^{1-\gamma}}{d^q}$,

for some C_1, C_2, C_3 that depend only on p. ∎

Remark: The proof holds with minor modifications when $n_i = \lfloor \lg(ci) \rfloor$ for any constant c. However note that if $c \geq 4$ then $n_1 > 1$, so the wall is dominated (by Proposition 2.4).

We can now proceed to prove that "thin" walls (i.e., $n \leq \lfloor \lg 2i \rfloor$) have Condorcet failure probabilities. We cannot expect all the walls with sub-logarithmic row widths to have a Condorcet F_p, since by Proposition 4.2, F_p of a bounded width wall is not Condorcet. Instead we prove that F_p is Condorcet for any wall in which each width k appears in at least 2^{k-1} consecutive rows, and the width increases by at most 1 from row to row. Clearly these walls have sub-logarithmic row widths. To prove this result (Theorem 6.4) we need some definitions and a technical lemma.

Since we are only interested in sequences (n_i) that grow more slowly than $n_i = i$, we can assume that n_i goes over all the integers, by increments of at most 1. Therefore the sequence (n_i) can be grouped into blocks B_j such that $n_i = j$ for all i in block B_j.

Instead of looking at $\mathrm{CW}\langle \boldsymbol{n}^{[d]} \rangle$ for all values of d, we restrict our attention first to the subsequence of walls in which the last block of widths is full, i.e., $n_d = k$ but already $n_{d+1} = k+1$ for some k.

Definition 6.2 *Let (n_i) be a standard sequence with $n_i \leq n_{i+1} \leq n_i + 1$ for all i. Let m_j denote the length of the j'th block (i.e., the number of times the value j appears in the sequence). Let $d_k = \sum_{j=1}^{k} m_j$ be the length of the sequence prefix till the end of the k'th block. Let*

$$F(d_k) = F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d_k]} \rangle) = F_p(\mathrm{CW}\langle 1, 2, 2 \ldots, 3, 3 \ldots, k, k, \ldots, k \rangle)$$

denote the failure probability of the wall ending with a complete last block (i.e., the last m_k rows are of width k).

The following lemma compares the failure probabilities of walls at the ends of corresponding blocks.

Lemma 6.3 *Let (n_i) and (n'_i) be two sequences as in Definition 6.2, with block lengths m_j (m'_j resp.), and failure probabilities at block ends $F(d_k)$ ($F'(d'_k)$ resp.). If $m'_j \geq m_j$ for $1 \leq j \leq k$ then $F'(d'_k) \leq F(d_k)$ for all $0 < p < \frac{1}{2}$.*

Proof: We use induction on the block number, k. For the induction base, note that by the definition $d_1 = d'_1 = 1$ so $F'(d'_1) = F(d_1) = p$.

Assume the claim holds up to $k-1$. Note that the last m'_k rows of $\mathrm{CW}\langle \boldsymbol{n}^{[d'_k]} \rangle$ are of width k. Therefore by applying the recurrence of Fact 2.11 m'_k times, and using the formula for a finite geometric series, we get that

$$F'(d'_k) = (1 - p^k - q^k)^{m'_k} F'(d'_{k-1}) + \frac{p^k}{p^k + q^k}[1 - (1 - p^k - q^k)^{m'_k}]. \tag{5}$$

Let $\gamma = (1 - p^k - q^k)$. We need to show that

$$\gamma^{m'_k} F'(d'_{k-1}) + \frac{p^k}{p^k + q^k}(1 - \gamma^{m'_k}) \leq F(d_k).$$

Using the induction hypothesis on the left-hand side and expanding the right-hand side in the same manner as in (5), it suffices to show that

$$\gamma^{m'_k} F(d_{k-1}) + \frac{p^k}{p^k + q^k}(1 - \gamma^{m'_k}) \leq \gamma^{m_k} F(d_{k-1}) + \frac{p^k}{p^k + q^k}(1 - \gamma^{m_k}),$$

which is equivalent to showing that

$$\frac{p^k}{p^k+q^k}(\gamma^{m_k}-\gamma^{m'_k}) \le F(d_{k-1})(\gamma^{m_k}-\gamma^{m'_k}). \tag{6}$$

By the premise $m'_k \ge m_k$. If $m'_k = m_k$, we are done. Otherwise $\gamma^{m_k}-\gamma^{m'_k} > 0$, so (6) turns into $F(d_{k-1}) \ge p^k/(p^k+q^k)$. This holds by Lemma 3.4, since d_{k-1} is the last row with width $k-1$. ∎

Now we can proceed to prove that slow-growth crumbling wall families have failure probabilities that are Condorcet.

Theorem 6.4 *Let (n_i) be a sequence as in Definition 6.2, with block lengths m_j and block ends at d_k. If $m_j \ge 2^{j-1}$ for all j then $F(d) = F_p(\mathrm{CW}\langle \boldsymbol{n}^{[d]}\rangle) \underset{d\to\infty}{\longrightarrow} 0$ for $0 < p < \frac{1}{2}$, thus $\mathrm{CW}\langle \boldsymbol{n}^{[d]}\rangle$ has a Condorcet failure probability.*

Proof: Let $(\hat{n}_i)$ denote the CWlog wall, i.e., $\hat{n}_i = \lfloor \lg 2i \rfloor$, with block lengths $\hat{m}_j$ and block ends at $\hat{d}_k$. Let $\hat{F}(d) = F_p(\mathrm{CW}\langle \hat{\boldsymbol{n}}^{[d]}\rangle)$ denote the failure probability of CWlog. Clearly $\hat{m}_j = 2^{j-1}$.

Consider $F(d)$ at block ends, in comparison to CWlog. By the premise, $m_j \ge \hat{m}_j$ for all j. Therefore by Lemma 6.3, $F(d_k) \le \hat{F}(\hat{d}_k)$ for all k. By Theorem 6.1 $\hat{F}(d) \underset{d\to\infty}{\longrightarrow} 0$ for all $0 < p < \frac{1}{2}$, so $F(d_k) \underset{k\to\infty}{\longrightarrow} 0$ as well. So far we have shown that $F(d)$ has a vanishing subsequence. However $F(d)$ is bounded and strictly decreasing by Proposition 3.5, so it has a unique limit, i.e., $\lim_{d\to\infty} F(d) = \lim_{k\to\infty} F(d_k) = 0$. ∎

Remarks:

- The requirement $m_j \ge 2^{j-1}$ for all j implies that the row width n_i is unbounded (otherwise, if $n_i \le k'$ for all i then the sequence never reaches the value $k'+1$ so $m_{k'+1} = 0$). This is as expected in view of Proposition 4.2.
- It is easy to see that any sequence (n_i) that fills the conditions of Theorem 6.4 is "thinner" than CWlog, i.e., $n_i \le \lfloor \lg 2i \rfloor$ for all i.

Acknowledgment

We are grateful to Moni Naor for many stimulating discussions.

References

[AE91] D. Agrawal and A. El-Abbadi. An efficient and fault-tolerant solution for distributed mutual exclusion. *ACM Trans. Comp. Sys.*, 9(1):1–20, 1991.

[BG87] D. Barbara and H. Garcia-Molina. The reliability of vote mechanisms. *IEEE Trans. Comput.*, C-36:1197–1208, October 1987.

[CAA90] S. Y. Cheung, M. H. Ammar, and M. Ahamad. The grid protocol: A high performance scheme for maintaining replicated data. In *Proc. 6th IEEE Int. Conf. Data Engineering*, pages 438–445, 1990.

[Con] N. Condorcet. Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralite des voix. Paris, 1785.

[DGS85] S. B. Davidson, H. Garcia-Molina, and D. Skeen. Consistency in partitioned networks. *ACM Computing Surveys*, 17(3):341–370, 1985.

[DKK+94] K. Diks, E. Kranakis, D. Krizanc, B. Mans, and A. Pelc. Optimal coteries and voting schemes. *Inf. Proc. Letters*, 51:1–6, 1994.

[EL75] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and Finite Sets*, pages 609–627. Colloq. Math. Soc. János Bolyai 10, 1975.

[GB85] H. Garcia-Molina and D. Barbara. How to assign votes in a distributed system. *J. ACM*, 32(4):841–860, 1985.

[Gif79] D. K. Gifford. Weighted voting for replicated data. In *Proc. 7th Symp. Oper. Sys. Princip.*, pages 150–159, 1979.

[Her84] M. P. Herlihy. *Replication Methods for Abstract Data Types.* PhD thesis, Massachusetts Institute of Technology, MIT/LCS/TR-319, 1984.

[HMP95] R. Holzman, Y. Marcus, and D. Peleg. Load balancing in quorum systems. In *Proc. 4th Workshop on Algorithms and Data Structures*, Kingston, Ont., Canada, 1995. To appear in SIAM J. Discrete Math.

[KC91] A. Kumar and S. Y. Cheung. A high availability $\sqrt{n}$ hierarchical grid algorithm for replicated data. *Inf. Proc. Letters*, 40:311–316, 1991.

[KRS93] A. Kumar, M. Rabinovich, and R. K. Sinha. A performance study of general grid structures for replicated data. In *Proc. Inter. Conf. Dist. Comp. Sys.*, 1993.

[Kum91] A. Kumar. Hierarchical quorum consensus: A new algorithm for managing replicated data. *IEEE Trans. Comput.*, 40(9):996–1004, 1991.

[Lov73] L. Lovász. Coverings and colorings of hypergraphs. In *Proc. 4th Southeastern Conf. Combinatorics, Graph Theory and Computing*, pages 3–12, 1973.

[Mae85] M. Maekawa. A $\sqrt{n}$ algorithm for mutual exclusion in decentralized systems. *ACM Trans. Comp. Sys.*, 3(2):145–159, 1985.

[MP92] Y. Marcus and D. Peleg. Construction methods for quorum systems. Technical Report CS92–33, The Weizmann Institute of Science, Rehovot, Israel, 1992.

[MV88] S. J. Mullender and P. M. B. Vitányi. Distributed match-making. *Algorithmica*, 3:367–391, 1988.

[Nei92] M. L. Neilsen. *Quorum Structures in Distributed Systems.* PhD thesis, Dept. Computing and Information Sciences, Kansas State University, 1992.

[NW94] M. Naor and A. Wool. The load, capacity and availability of quorum systems. In *Proc. 35th IEEE Symp. Found. of Comp. Science*, pages 214–225, 1994.

[NW96] M. Naor and A. Wool. Access control and signatures via quorum secret sharing. In *Proc. 3rd ACM Conf. Comp. and Comm. Security*, New Delhi, India, 1996. To appear, see also Technical Report CS95-19, The Weizmann Institute of Science.

[PW95a] D. Peleg and A. Wool. The availability of quorum systems. *Information and Computation*, 123(2):210–223, 1995.

[PW95b] D. Peleg and A. Wool. Crumbling walls: A class of practical and efficient quorum systems. In *Proc. 14th ACM Symp. Princip. of Dist. Comp.*, pages 120–129, Ottawa, Canada, 1995.

[Ray86] M. Raynal. *Algorithms for Mutual Exclusion.* MIT press, 1986.

[RST92] S. Rangarajan, S. Setia, and S. K. Tripathi. A fault-tolerant algorithm for replicated data management. In *Proc. 8th IEEE Int. Conf. Data Engineering*, pages 230–237, 1992.

[RT91] S. Rangarajan and S. K. Tripathi. A robust distributed mutual exclusion algorithm. In *Proc. 5th Inter. Workshop on Dist. Algorithms, LNCS 579*, pages 295–308. Springer-Verlag, 1991.

[SB94] M. Spasojevic and P. Berman. Voting as the optimal static pessimistic scheme for managing replicated data. *IEEE Trans. Par. Dist. Sys.*, 5(1):64–73, 1994.

[Tho79] R. H. Thomas. A majority concensus approach to concurrency control for multiple copy databases. *ACM Trans. Database Sys.*, 4(2):180–209, 1979.

[YG94] T. W. Yan and H. Garcia-Molina. Distributed selective dissemination of information. In *Proc. 3rd Inter. Conf. Par. Dist. Info. Sys.*, pages 89–98, 1994.