Analysis of Quorum-Based Protocols for Distributed (k+1)-Exclusion

Divyakant Agrawal, Ömer Eğecioğlu, and Amr El Abbadi

Department of Computer Science, University of California
Santa Barbara, CA 93106, USA

Abstract. A generalization of the majority quorum for the solution of the distributed $(k+1)$-exclusion problem is proposed. This scheme produces a family of quorums of varying sizes and availabilities indexed by integral divisors r of k. The cases $r = 1$ and $r = k$ correspond to known majority based quorum generation algorithms MAJ and DIV, whereas intermediate values of r interpolate between these two extremes. A cost and availability analysis of the proposed methods is also presented. An interesting result of this analysis is that in a reasonably reliable environment with a large number of sites, even protocols with low communication costs attain high availability.

1 Introduction

The problem of distributed mutual exclusion has been extensively studied and many interesting protocols for its solution have been proposed. Most of these protocols attempt to provide high performance by reducing the number of messages involved or by improving the degree of fault-tolerance and hence improving the chances of achieving mutual exclusion in the presence of site and communication failures. A generalization of the mutual exclusion problem is the *k-mutual exclusion problem*, where no more than k processes are allowed to enter the critical section simultaneously. Since $k+1$ or more processes are excluded, this problem is also referred to as the *$(k+1)$-exclusion problem*.

In a distributed environment, the $(k+1)$-exclusion problem arises in several interesting applications. For example, it could be used to monitor the number of processes in a distributed system that are allowed to perform a certain action, such as issuing broadcast messages. In such a case, the system may restrict the number of broadcasting processes so as to control the level of congestion. Another application is in the context of replicated databases that allow *bounded ignorance* [10], i.e., when transactions may specify that they do not need to be aware of the k most recent updates to the database. Here also, instead of the traditional database system that uses distributed mutual exclusion to ensure one update to the replicated data at any time, several updates may be permitted simultaneously. Efficient and highly available solutions to the distributed $(k+1)$-exclusion problem would be particularly useful for such applications.

The distributed $(k+1)$-exclusion problem was first solved by Raymond [14], who provided a simple extension to the Ricart and Agrawala's mutual exclusion algorithm [15]. Srimani and Reddy [18] improved on this protocol by using the notion of *privilege* of Suzuki and Kasami [19]. This solution reduces the number of required messages to achieve mutual exclusion. Recently, there has been a significant interest in fault-tolerant methods to solve the $(k+1)$-exclusion problem based on the notion of quorums [7]. Fujita et al. [5] discuss some simple techniques and then propose a scheme with small quorum sizes. Huang et al. [8] propose an alternative method with small quorums and high availability in the presence of failures.

The majority quorum algorithm [20, 7] for distributed mutual exclusion has been widely used to develop various quorum-based protocols for mutual exclusion as well as for $(k+1)$-exclusion by a suitable partitioning of the sites. Agrawal and El Abbadi [1] partition the sites in a network to construct majority based quorums defined on hybrid logical structures such as a grid [12] and a tree [2]. Rangarajan and Tripathi [13] partition the sites into N classes and organize the classes to form a finite projective plane. Quorums are then defined using $\sqrt{N}$ classes and within each class a majority of sites is chosen. Kumar [11] uses the partitioning approach to recursively define hierarchical quorums based on the majority rule and quorums. Fujita et al. [5] also use a partitioning approach for $(k+1)$-exclusion in which the sites are partitioned into k clusters and quorums are constructed so that mutual exclusion is ensured within each cluster.

In this paper we generalize the majority quorum algorithm for constructing quorums for $(k+1)$-exclusion and analyze the resulting methods. We propose a sequence of algorithms MAJ_r with varying costs and availabilities, indexed by integral divisors r of k. When $r = 1$, the quorums correspond to those produced by MAJ (which partitions the sites into a single class), and when $r = k$, the quorums correspond to the quorums produced by DIV (which partitions the sites into k classes) [5]. For intermediate values of r, a sliding tradeoff between the number of quorum sets and quorum size is achieved.

Our analysis of availability is based on estimates for truncated binomial sums [3, 13], and the Vandermonde convolution identity. Recently, Kakugawa et al. [9] analyzed MAJ versus a centralized site solution adapted for $(k+1)$-exclusion called SGL, and showed that for networks with low probability of site failure, MAJ provides optimal availability performance over *all* quorum-based mechanisms, whereas for sites with a high probability of failures SGL gives higher availability than MAJ. Our analysis is similar except that our quorums generalize majority, and instead of the two extremes, we consider a whole spectrum of protocols. The analytical results we present are of two kinds:

1. For p close to zero (where p is the probability that a site is up in a system of n sites), we show that there is a nonzero p_n such that for $p < p_n$, DIV has better availability than MAJ. Furthermore the limit of p_n as n gets large is nonzero.
2. For p close to 1, the increase in availability achieved by MAJ for large n is insignificant compared to that attained by MAJ_r for $r > 1$.

Note that 1 is mostly of theoretical interest, whereas 2 has practical implications for systems with low probability of site failures. This is because the size of the quorums produced by MAJ_r for $r > 1$ are as small as a half of those produced by MAJ, and consequently the communication costs of acquiring such quorums is less than the cost of acquiring a MAJ quorum. Thus for highly available geographically dispersed distributed systems with a large number of sites, MAJ_r for $r > 1$ may be preferable over MAJ. Even when individual site availability is greater than about 75%, the increased availability offered by MAJ becomes inconsequential when compared even to DIV, which falls on the other end of the spectrum. Intermediate values of r serve to further interpolate between communication cost and availability.

2 The Problem Statement

A distributed system consists of a set of sites that communicate with each other by sending messages over a communication network. We assume that every site has the capability to send a message to any other site when there is a communication path between them. The sites are either *fail-stop* or may fail to send or receive messages. Communication links may fail by crashing, or by failing to deliver messages. Although quorum-based protocols are resilient both site and communication failures, our analysis assumes site failures only [13, 9].

Distributed mutual exclusion is a classical technique for providing access to shared resources. We postulate the existence of a resource in the network, which may be accessed by a single process at a time. To access the resource, a process (site) p_i is required to receive permission from a set of sites S_i. If all sites in S_i grant permission to p_i, then it is allowed to access the resource. To ensure mutual exclusion the sets S_i are required to satisfy the *intersection property*: For any $i \neq j$, $S_i \cap S_j \neq \phi$. These and related concepts were formalized and analyzed in terms of the notions of *quorums* and *coteries* [7, 12, 6]. In $(k+1)$-exclusion, up to k processes are allowed to access the resource simultaneously. Thus, if we consider $k+1$ sets of sites that grant permission to access the resource then there must exist at least two among these $k+1$ sets with a non-empty intersection. The $(k+1)$-exclusion problem can now be stated in terms of the requirements **(I)** and **(II)** below:

(I) The (k+1)-Intersection Property. For any $k+1$ sets $S_1, S_2, \cdots, S_{k+1}$, there exist two distinct sets S_i and S_j such that $S_i \cap S_j \neq \phi$.

Note that quorums constructed to ensure the traditional mutual exclusion condition also ensure the above property. Hence, in order to eliminate trivial solutions to the $(k+1)$-exclusion problem we add an additional restriction [5, 8].

(II) The k-Non-intersection Property. There exist k sets $S_1, S_2, \cdots, S_k$ such that for any two distinct sets S_i and S_j, $S_i \cap S_j = \phi$.

The second property above is desirable for all values of k. When $k = 1$, i.e., in the case of mutual exclusion, it is satisfied vacuously.

3 A General Paradigm for $(k+1)$-Exclusion

One of the simplest approaches to ensure mutual exclusion in a distributed system is to use majority quorums [7, 20] of size $\lfloor \frac{n}{2} \rfloor + 1$. For 3-exclusion, we can reduce the size of the permission sets to $\lfloor \frac{n}{3} \rfloor + 1$. Clearly, the 3-intersection property holds since any three sets of sites with size $\lfloor \frac{n}{3} \rfloor + 1$ chosen from n sites will always have two sets with non-empty intersection. Similarly, the 2-non-intersection property also holds for $n > 5$, since it is possible to construct two disjoint sets when only one-third the number of sites from n are used for each. For $(k+1)$-exclusion, it suffices to take the quorum size to be $\lfloor \frac{n}{k+1} \rfloor + 1$. This majority based construction for the $(k+1)$-exclusion problem is referred to as MAJ [5].

Another approach for achieving $(k+1)$-exclusion is to consider k instances of any mutual exclusion solution. A process wishing $(k+1)$ exclusive access to a resource acquires permission from any of the k instances. This ensures the $(k+1)$-intersection property since any $k+1$ quorums chosen will consist of at least two quorums in the same instance of the mutual exclusion solution, and hence must have a non-empty intersection. Similarly, the k-non-intersection property is satisfied if each of the k processes choose a quorum from different instances of the mutual exclusion solution. This construction for the $(k+1)$-exclusion problem is referred to as DIV [5].

The two generalizations MAJ and DIV of majority quorums for the solution of the $(k+1)$-exclusion problem are at the opposite ends of a spectrum. In MAJ, the original mutual exclusion majority solution is generalized whereas in DIV the sites in the network are partitioned into k classes with each class using any traditional approach to enforce mutual exclusion. We explore the possibility of enforcing $(k+1)$-exclusion by varying the number r of classes from 1 to k, and define a quorum generation method MAJ_r for any r dividing k. To simplify the presentation, we assume we are given a set of n sites where $n = kN$ for some $N \geq 1$. In MAJ_r, the $(k+1)$-exclusion problem is solved by partitioning the sites into r disjoint classes where $r = k/i$, for some integer i. Within each class, we choose the quorums of size q_r which guarantee that at least two sets from any collection of $i+1$ sets within the class intersect. More precisely MAJ_r denotes the method in which

1. The sites $1, 2, \ldots, n$ are partitioned into r classes of size $n/r = iN$ each.
2. From each class, all subsets of size

$$q_r = \left\lfloor \frac{iN}{i+1} \right\rfloor + 1 \ = \ \left\lfloor \frac{n}{k+r} \right\rfloor + 1 \tag{1}$$

   are taken as quorums. Here $r = k/i$.

It should be clear that MAJ_r produces sets that satisfy both **(I)** and **(II)**, aside from the trivial cases of small parameters for which $\lfloor \frac{n}{k+r} \rfloor = \lfloor \frac{n}{k+r-1} \rfloor$ (see [21]). Furthermore, $MAJ_1 = MAJ$, and $MAJ_k = DIV$ are special cases of this construction.

4 Availability Measures

The communication cost associated with obtaining mutual exclusion by using the quorum approach is directly proportional to the quorum size. The availability and the fault-tolerance characteristics of a particular method are determined by the number of ways in which a quorum can be constructed from a given number of sites in the network. The total number of sets produced by MAJ_r is

$$r\binom{iN}{q_r}, \tag{2}$$

where q_r is as given in (1). Let $q_1 = q_{MAJ}$ be the quorum size and T_{MAJ} be the number of sets in MAJ. Then

$$q_{MAJ} = \left\lfloor \frac{n}{k+1} \right\rfloor + 1\ , \qquad T_{MAJ} = \binom{n}{q_{MAJ}}\ . \tag{3}$$

The other extreme of the partitioning approach arises when $r = k$. The quorum size $q_k = q_{DIV}$ and the total number of quorums T_{DIV} for DIV are given by the formulas

$$q_{DIV} = \left\lfloor \frac{n}{2k} \right\rfloor + 1\ , \qquad T_{DIV} = k\binom{\frac{n}{k}}{q_{DIV}}\ . \tag{4}$$

When MAJ and DIV are evaluated purely in terms of the communication costs incurred to enforce $(k+1)$-exclusion, DIV is preferable due to its smaller sized quorums. However, if the evaluation criterion includes the number of quorum sets produced, the outcome is not so trivial. Fujita et al. [5] conjectured that the partitioned approach restricts the number of ways a quorum can be selected and hence will provide inferior availability. First we explore this issue in the context of these two approaches and isolate the instances in which DIV actually performs better than MAJ.

Suppose p is the probability that a site is up. If $q = q_r$ is the quorum size given by the formula (1), then the probability $C_r(p)$ that a quorum set is available in a given class for the method MAJ_r is a polynomial $C_r(p)$ of degree iN in p (recall that there are a total of r classes of iN sites each):

$$C_r(p) = \binom{iN}{q} p^q (1-p)^{iN-q} + \binom{iN}{q+1} p^{q+1} (1-p)^{iN-q-1} + \cdots + \binom{iN}{iN} p^{iN}\ . \tag{5}$$

Let $AV_r(p)$ denote the probability that a quorum set is available when the method MAJ_r is used. For the extreme cases of MAJ and DIV, we also use the notation $AV_{MAJ}(p)$ for $AV_1(p)$, and $AV_{DIV}(p)$ for $AV_k(p)$. The probability that none of the r classes of the partition has a quorum set available is $(1 - C_r(p))^r$, and therefore

$$AV_r(p) = 1 - (1 - C_r(p))^r\ . \tag{6}$$

Example 1. Suppose $k = 2$ and $n = 10$. Then for $i = 2$ we get $r = 1$ and $MAJ_1 = MAJ$. From (5),

$$C_{MAJ}(p) = 210p^4(1-p)^6 + 252p^5(1-p)^5 + 210p^6(1-p)^4 + 120p^7(1-p)^3 + \\ + 45p^8(1-p)^2 + 10p^9(1-p) + p^{10}.$$

For $i = 1$, $r = 2$ and $MAJ_2 = DIV$. In this case $C_{DIV}(p) = 10p^3(1-p)^2 + 5p^4(1-p) + p^5$. Therefore, from (6)

$$AV_{MAJ}(p) = C_{MAJ}(p),$$
$$AV_{DIV}(p) = 20p^3 - 30p^4 + 12p^5 - 100p^6 + 300p^7 - 345p^8 + 180p^9 - 36p^{10}\ .$$

Since $AV_{DIV}(p) - AV_{MAJ}(p) = 20p^3(1-p)^6(1-6p)$, setting $AV_{DIV}(p) > AV_{MAJ}(p)$ and solving for p, we find that whenever $p < 1/6$, DIV provides better quorum availability than MAJ.

5 Availability of MAJ and DIV for 3-Exclusion

We now generalize the above example by deriving asymptotic results for the availabilities of MAJ_r for $r = 1$ (MAJ) and $r = k$ (DIV). For simplicity of exposition, we start by presenting proofs for the case $k = 2$ (3-exclusion). In the appendix, we provide the proofs for arbitrary k. We start out by studying the polynomials $AV_{MAJ}(p)$ and $AV_{DIV}(p)$ for large n. Since in this case

$$q_1 = q_{MAJ} = \left\lfloor \frac{n}{3} \right\rfloor + 1 \ , \quad q_2 = q_{DIV} = \left\lfloor \frac{n}{4} \right\rfloor + 1$$

from formula (1), it is convenient to assume that $n = 12m$ for some $m \geq 1$. Then $q_{MAJ} = 4m+1, \ \ q_{DIV} = 3m+1$ with

$$AV_{MAJ}(p) = \sum_{j=4m+1}^{12m} \binom{12m}{j} p^j (1-p)^{12m-j} \ ,$$

$$AV_{DIV}(p) = 1 - \left[1 - \sum_{j=3m+1}^{6m} \binom{6m}{j} p^j (1-p)^{6m-j} \right]^2 \ .$$

The values of p in $0 < p < p_n$ for which $AV_{DIV}(p) > AV_{MAJ}(p)$ for small values of $n = 12m$ are tabulated in Table 1 (computed with the aid of MACSYMA). Even though the numbers p_n are decreasing, the limit of p_n as n gets large is nonzero, as shown in Theorem 1.

n	12	24	36	48	60	72	84	96
p_n	0.0461	0.0423	0.0408	0.0399	0.0394	0.0391	0.0388	0.0386

Table 1. Computed range of values $0 < p < p_n$ for which $AV_{DIV}(p) > AV_{MAJ}(p)$.

Theorem 1. *For 3-exclusion, DIV provides better availability for large n than MAJ when the probability p of a site being up satisfies* $p < 0.0299$.

Proof. We set $AV_{DIV}(p) > AV_{MAJ}(p)$ and find the values of p that satisfy this inequality as n gets large. Since

$$1 - AV_{MAJ}(p) = \sum_{j=0}^{4m} \binom{12m}{j} p^j (1-p)^{12m-j} \ ,$$

$$1 - C_{DIV}(p) = \sum_{j=0}^{3m} \binom{6m}{j} p^j (1-p)^{6m-j} \ ,$$

$AV_{DIV}(p) > AV_{MAJ}(p)$ if and only if

$$\left[\sum_{j=0}^{3m} \binom{6m}{j} p^j (1-p)^{6m-j} \right]^2 < \sum_{j=0}^{4m} \binom{12m}{j} p^j (1-p)^{12m-j} \tag{7}$$

Let $q = p/(1-p)$. Factoring out and cancelling $(1-p)^{12m}$ from both sides, inequality (7) can be written in terms of q as

$$\left[\sum_{j=0}^{3m} \binom{6m}{j} q^j \right]^2 < \sum_{j=0}^{4m} \binom{12m}{j} q^j \ . \tag{8}$$

Write

$$\left[\sum_{j=0}^{3m}\binom{6m}{j}q^j\right]^2 = \sum_{j=0}^{6m}c_jq^j .$$

For $j \le 3m$,

$$c_j = \sum_{i=0}^{j}\binom{6m}{i}\binom{6m}{j-i} = \binom{12m}{j}$$

by the Vandermonde convolution identity [4, 16]. This means that the terms involving powers q^j up to $j = 3m$ on both sides of (8) are identical. Cancelling these terms, (8) can be written as

$$\sum_{j=4m+1}^{6m} c_jq^j < \sum_{j=3m+1}^{4m}\left[\binom{12m}{j} - c_j\right]q^j .$$

For large m and q close to zero, we can discard the effect of all but the lowest power of q on both sides. Thus for large n, in order for the inequality $AV_{DIV}(p) > AV_{MAJ}(p)$ to hold, the parameter q should satisfy

$$c_{4m+1}q^{4m+1} < \left[\binom{12m}{3m+1} - c_{3m+1}\right]q^{3m+1} .$$

However,

$$c_{3m+1} = \sum_{i=1}^{3m}\binom{6m}{i}\binom{6m}{3m+1-i} = \binom{12m}{3m+1} - 2\binom{6m}{3m+1}$$

by the Vandermonde identity again. Similarly,

$$c_{4m+1} = \sum_{i=m+1}^{3m}\binom{6m}{i}\binom{6m}{4m+1-i} = \binom{12m}{4m+1} - 2\sum_{i=0}^{m}\binom{6m}{i}\binom{6m}{3m+1-i}$$

In particular

$$c_{4m+1} < \binom{12m}{4m+1} .$$

Therefore, in order to guarantee that $AV_{DIV}(p) > AV_{MAJ}(p)$, it suffices to have

$$q^m < 2\binom{6m}{3m+1}/\binom{12m}{4m+1} .$$

Using Stirling's approximation,

$$\binom{am}{bm}^{1/m} \sim \frac{a^a}{b^b(a-b)^{a-b}} \tag{9}$$

for integers $a > b$. Therefore

$$\binom{12m}{4m+1} \sim \frac{3^{12m+1}}{2^{8m+1}\sqrt{3\pi m}} \quad \text{and} \quad \binom{6m}{3m+1} \sim \frac{2^{6m}}{\sqrt{3\pi m}} .$$

It follows that

$$q^m < (\frac{2^{6m+1}}{\sqrt{3\pi m}})/(\frac{3^{12m+1}}{2^{8m+1}\sqrt{3\pi m}}) = \frac{2^{14m}}{3^{12m+1}} .$$

Taking m-th roots and letting m go to infinity, we find that the condition $q < 2^{14}/3^{12}$ guarantees that $AV_{DIV}(p) \ge AV_{MAJ}(p)$. In terms of p this bound becomes

$$p < \frac{2^{14}/3^{12}}{1 + 2^{14}/3^{12}} = 0.029907...$$

which is the content of the Theorem. $\square$

The above result for MAJ and DIV for 3-exclusion can be generalized in two directions, both valid for large n. First of all, it is possible to analytically compare the availability of MAJ and DIV for arbitrary k and find the values of p for which DIV performs better than MAJ as a function of k. Alternately, given k, it is possible to compute the range of values of p for which MAJ_r performs better than MAJ. In the former setting there is a constant $b_k = O(1/k^2)$ such that whenever $0 < p < b_k$, $AV_{DIV}(p) > AV_{MAJ}(p)$ for large n (i.e. in this range of p, DIV provides better availability than MAJ for $(k+1)$-exclusion). This is formally established in Theorem 2. For the latter generalization, there is a function $B(k,r)$ such that for $p < B(k,r)$, $AV_r(p) > AV_{MAJ}(p)$ for large n. (i.e. in this range of p, MAJ_r provides better availability than MAJ for $(k+1)$-exclusion). This is shown in Theorem 3.

Theorem 2. *For large n, $AV_{DIV}(p) \geq AV_{MAJ}(p)$ for $0 < p < b_k$, where*

$$b_k \sim \frac{4}{4 + e^2(k+1)^2}\ , \tag{10}$$

and $e = 2.7182\ldots$.

Proof. See Appendix.

Theorem 3. *For large n and k, $AV_r(p) \geq AV_{MAJ}(p)$ whenever the probability p of a site being up satisfies*

$$p/(1-p) \ < \ \frac{1}{r^{k/(r-1)}(k+1)^{(k+r)/(r-1)}}\ .$$

Proof. See Appendix.

These generalizations are of theoretical interest as they are valid only for unrealistically small values of p.

6 Availability of MAJ_r Quorums

In general MAJ has better availability than DIV as well as all the other MAJ_r, except for systems with high site failure probability. On the other hand, from the formula in (1), MAJ quorums have twice the size of DIV, and are always larger than MAJ_r quorums for $r > 1$. In this section, we show that for large n, and p close to 1, the increase in availability provided by MAJ itself as compared to MAJ_r is actually quite small. This is significant since in most current systems we expect sites to have a low probability of failure. Hence in these cases we can use the smaller sized quorums of MAJ_r without losing much on availability, while reducing the communication overhead by up to a factor of 2.

Let $D_r(p) = |AV_{MAJ}(p) - AV_r(p)|$ denote the magnitude of the absolute error in availability made when MAJ_r instead of MAJ itself is used. Then

$$D_r(p) = \left| \sum_{j=0}^{q_{MAJ}-1} \binom{n}{j} p^j (1-p)^{n-j} - \left[\sum_{j=0}^{q_r-1} \binom{n/r}{j} p^j (1-p)^{n/r-j} \right]^r \right|$$

Factoring $(1-p)^n$ and letting $q = p/(1-p)$, we can write

$$D_r(p) = (1-p)^n \left| \sum_{j=0}^{q_{MAJ}-1} \binom{n}{j} q^j - \left[\sum_{j=0}^{q_r-1} \binom{n/r}{j} q^j \right]^r \right|$$

Consider the coefficients c_j in the expansion

$$\left[\sum_{j=0}^{q_r-1} \binom{n/r}{j} q^j \right]^r = \sum_{j=0}^{r(q_r-1)} c_j q^j\ .$$

By the Vandermonde convolution identity [4, 16], $c_j \le \binom{n}{j}$, with equality in the range $0 \le j \le q_r - 1$. Therefore

$$D_r(p) = (1-p)^n \left| \sum_{j=q_r}^{q_{MAJ}-1} \binom{n}{j} q^j - \sum_{j=q_r}^{r(q_r-1)} c_j q^j \right| \le (1-p)^n \sum_{j=q_r}^{r(q_r-1)} \binom{n}{j} q^j .$$

However, q^j is an increasing function of j whenever $p \ge 0.5$. Thus for $p \ge 0.5$

$$D_r(p) \le p^{r(q_r-1)}(1-p)^{n-r(q_r-1)} \sum_{j=q_r}^{r(q_r-1)} \binom{n}{j} .$$

But since $1 \le r \le k$, $r(q_r - 1) = (r/(k+r))\,n \le n/2$, and therefore

$$\sum_{j=q_r}^{r(q_r-1)} \binom{n}{j} \le 2^{n-1} .$$

Furthermore, $n - (r/(k+r))\,n = (k/(k+r))\,n$, and 1 is an upper bound for $p^{r(q_r-1)}$. Therefore

$$D_r(p) \le \frac{1}{2}\,[2(1-p)^{\frac{k}{k+r}}]^n .$$

If p is close enough to 1 so that $2(1-p)^{\frac{k}{k+r}} < 1$, or equivalently when

$$p > 1 - (\frac{1}{2})^{\frac{k+r}{k}} , \tag{11}$$

the error $D_r(p)$ goes to zero as n gets large. To summarize, we have proved

Theorem 4. *If p is in the range given by* (11), *then*

$$|AV_{MAJ}(p) - AV_r(p)| \le \frac{1}{2}\,[2(1-p)^{\frac{k}{k+r}}]^n , \tag{12}$$

and the right hand side of (12) *goes to zero as n gets large.*

To get a rough idea of the magnitude of the difference in availability which will hold for all algorithms MAJ_r at once, we note that $(k+r)/k \le 2$ since r is a divisor of k. Therefore whenever $p > 1 - (1/2)^2 = 0.75$, the error in availability satisfies

$$|AV_{MAJ}(p) - AV_r(p)| \le \frac{1}{2}\,(2\sqrt{1-p}\,)^n , \tag{13}$$

regardless of the value of r.

Example 2. Consider a system with $n = 100$ sites in which the probability of a site being up is $p = 0.9$. From (13) the difference between the availabilities of MAJ and DIV for any $(k+1)$-exclusion is about 6.3×10^{-21}, a negligible amount. On the other hand when $k = 9$, MAJ requires quorums of size $q_{MAJ} = 11$, while DIV requires quorums of size only $q_{DIV} = 6$.

7 Discussion

In this paper we proposed a family of quorum-based protocols MAJ_r that generalize majority quorums for distributed $(k+1)$-exclusion. These protocols are indexed by integral divisors r of k, with $MAJ_1 = MAJ$ and $MAJ_k = DIV$. In addition, we analyzed the whole spectrum of resulting protocols with respect to availability and communication cost. We first showed that each MAJ_r, $r > 1$, provides higher availability than MAJ in a large network where the probability of a site being up is sufficiently small. Next, we concentrated on the other end of the probability spectrum in which the network has a low probability of site failure. In this case, we showed that the increase in availability achieved by MAJ over any member of the family MAJ_r, $r > 1$ decreases rapidly as the number of sites gets large. Since the communication overhead of solving the $(k+1)$-exclusion problem is directly proportional to the quorum sizes, reductions up to a factor of two in communication overhead can be achieved without significant sacrifice in availability.

References

1. D. Agrawal and A. El Abbadi. Exploiting Logical Structures of Replicated Databases. *IPL*, 33(5):255–260, Jan. 1990.
2. D. Agrawal and A. El Abbadi. An Efficient and Fault-tolerant Solution for Distributed Mutual Exclusion. *ACM Trans. on Computer Sys.*, pp. 1–20, Feb. 1991.
3. M. Ahamad and M. H. Ammar. Performance Characterization of Quorum-Consensus Algorithms for Replicated Data. *IEEE Trans. on Software Engineering*, 15(4):492–495, Apr. 1989.
4. D. Cohen. *Basic Techniques of Combinatorial Theory.* John Wiley & Sons, 1978.
5. S. Fujita, M. Yamashita, and T. Ae. Distributed k-mutual exclusion problem and k-coteries. In *Proc. on the Symp. on Algorithms*, pp. 22–31, 1991.
6. H. Garcia-Molina and D. Barbara. How to assign votes in a distributed system. *J. of the ACM*, 32(4):841–860, Oct. 1985.
7. D. K. Gifford. Weighted Voting for Replicated Data. In *Proc. of 7th ACM Symp. on Operating Systems Principles*, pp. 150–159, Dec. 1979.
8. S. Huang, J. Jiang, and Y. Kuo. K-coteries for fault-tolerant k entries to a critical section. In *Proc. of 13th Int. Conf. on Dist. Comp. Sys.*, pp. 74–81, June 1993.
9. H. Kakugawa, S. Fujita, M. Yamashita, and T. Ae. Availability of k-Coteries. *IEEE Trans. on Computers*, 42(5):553–558, May 1993.
10. N. Krishnakumar and A. J. Bernstein. Bounded Ignorance in Replicated Systems. In *Proc. of 10th ACM Symp. on Prin. of Database Sys.*, pp. 63–74, May 1991.
11. A. Kumar. Hierarchical Quorum Consensus - A New Algorithm for Managing Replicated Data. *IEEE Trans. on Computers*, 39(9):996–1004, Sep. 1991.
12. M. Maekawa. A $\sqrt{n}$ algorithm for mutual exclusion in decentralized systems. *ACM Trans. on Computer Systems*, 3(2):145–159, May 1985.
13. S. Ranagarajan and S. Tripathi. Robust Distributed Mutual Exclusion Algorithms. In *Proc. of 5th Int. Workshop on Distributed Algorithms*, pp. 295–308, Oct. 1991. Lecture Notes in Computer Science, 579, Springer-Verlag.
14. K. Raymond. A distributed algorithm for multiple entries to a critical section. *IPL*, 30(4):189–193, Feb. 1989.
15. G. Ricart and A. K. Agrawala. An optimal algorithm for mutual exclusion in computer networks. *Communications of the ACM*, 24(1):9–17, January 1981.
16. J. Riordan. *An Introduction to Combinatorial Analysis.* Princeton Univ. Press, 1978.
17. R. Schlichting and F. B. Schneider. Fail-Stop Processors: An Approach to Designing Fault-Tolerant Computing Systems. *ACM Trans. on Computer Systems*, 1(3):222–238, Aug. 1982.
18. P. K. Srimani and R. L. N. Reddy. Another distributed algorithm for multiple entries to a critical section. *IPL*, 41(1):51–57, Jan. 1992.
19. I. Suzuki and T. Kasami. A Distributed Mutual Exclusion Algorithm. *ACM Trans. on Computer Systems*, 3(4):344–349, Nov. 1985.
20. R. H. Thomas. A Majority Consensus Approach to Concurrency Control for Multiple Copy Databases. *ACM Trans. on Database Systems*, 4(2):180–209, June 1979.
21. S.-M. Yuan and H.-K. Chang. Comments on "Availability of k-Coterie". *IEEE Trans. on Computers*, 43(12):1457, Dec. 1994.

8 Appendix

In this appendix, we include proofs that generalize Theorem 1 from 3-exclusion to $(k+1)$-exclusion for arbitrary k.

Theorem 2. *For large n, $AV_{DIV}(p) \geq AV_{MAJ}(p)$ for $0 < p < b_k$, where*

$$b_k \sim \frac{4}{4 + e^2(k+1)^2} \, ,$$

and $e = 2.7182\ldots$.

Proof. For the sake of simplicity of the analysis it is convenient to assume that n is of the form $n = 2k(k+1)m$ for some $m \geq 1$. By (3) and (4), the quorum sizes are of the form

$$q_{MAJ} = 2km + 1 \ , \qquad q_{DIV} = (k+1)m + 1 \ ,$$

for MAJ and DIV, respectively. In this case

$$AV_{MAJ}(p) = \sum_{j=2km+1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \tag{14}$$

$$AV_{DIV}(p) = 1 - \left[1 - \sum_{j=(k+1)m+1}^{n/k} \binom{n/k}{j} p^j (1-p)^{n/k-j}\right]^k \tag{15}$$

$AV_{DIV}(p) > AV_{MAJ}(p)$ if and only if

$$\left[\sum_{j=0}^{(k+1)m} \binom{n/k}{j} p^j (1-p)^{n/k-j}\right]^k < \sum_{j=0}^{2km} \binom{n}{j} p^j (1-p)^{n-j} \tag{16}$$

As before, let $q = p/(1-p)$. Then the inequality (16) can be written as

$$\left[\sum_{j=0}^{(k+1)m} \binom{n/k}{j} q^j\right]^k < \sum_{j=0}^{2km} \binom{n}{j} q^j \tag{17}$$

Write

$$\left[\sum_{j=0}^{(k+1)m} \binom{n/k}{j} q^j\right]^k = \sum_{j=0}^{n/2} c_j q^j \ .$$

For $j \leq (k+1)m$,

$$c_j = \sum_{i_1+i_2+\cdots+i_k=j} \binom{n/k}{i_1}\binom{n/k}{i_2}\cdots\binom{n/k}{i_k} = \binom{n}{j}$$

by the Vandermonde convolution identity. This means that the terms involving powers q^j up to $j = (k+1)m$ on both sides of (17) are identical. Therefore (17) can be written as

$$\sum_{j=2km+1}^{n/2} c_j q^j < \sum_{j=(k+1)m+1}^{2km} \left[\binom{n}{j} - c_j\right] q^j \ .$$

For large m and q close to zero, we can discard the effect of all but the lowest power of q on both sides. This means that for large n, the parameter q should satisfy

$$c_{2km+1} q^{2km+1} < \left[\binom{n}{(k+1)m+1} - c_{(k+1)m+1}\right] q^{(k+1)m+1} \ , \tag{18}$$

and therefore

$$q^{(k-1)m} < \left[\binom{n}{(k+1)m+1} - c_{(k+1)m+1}\right] / \ c_{2km+1} \ .$$

But

$$\binom{n}{(k+1)m+1} - c_{(k+1)m+1} = k\binom{n/k}{(k+1)m+1}$$

and

$$c_{2km+1} \ < \ \binom{n}{2km+1}$$

just as in the case of $k = 2$. By an analysis similar to the $k = 2$ case using (9), we find that in order to guarantee $AV_{DIV}(p) \geq AV_{MAJ}(p)$, it suffices to have for large n

$$q^{k-1} \ < \ [2^{2(k+1)} k^{2k^2}]/[(k+1)^{2k(k+1)}] \tag{19}$$

for large n. Taking the $(k-1)$-st root of the right hand side of (19) and using the fact that $e = \lim_{k\to\infty}(1+\frac{1}{k})^k$, it follows that asymptotically

$$q < \frac{4}{e^2(k+1)^2} \ . \tag{20}$$

Since $q = p/(1-p)$ implies $p = q/(1+q)$, (20) gives the stated value of the limit b_k of the theorem.

k	b_k	Approx. value
2	0.0299	0.0567
3	0.0188	0.0327
4	0.0127	0.0212
5	0.0093	0.0148
6	0.0071	0.0109
7	0.0056	0.0084
8	0.0045	0.0066
9	0.0037	0.0054
10	0.0032	0.0044

Table 2. Upper bound b_k for p for large n and k for which $AV_{DIV}(p) \geq AV_{MAJ}(p)$.

The values of the asymptotic formula on the right hand side of (10) for various k are given in the third column of Table 2. The second column lists the computed actual values of b_k. □

Theorem 3. *For large n and k, $AV_r(p) \geq AV_{MAJ}(p)$ whenever the probability p of a site being up satisfies*

$$p/(1-p) \ < \ \frac{1}{r^{k/(r-1)}(k+1)^{(k+r)/(r-1)}} \ .$$

Proof. Assume that n is of the form $n = r(k+1)(k+r)m$ for some $m \geq 1$ to simplify the analysis. Then

$$q_{MAJ} = r(k+r)m + 1 \ , \quad q_r = r(k+1)m + 1 \ .$$

From (5) and (6)

$$AV_1(p) = \sum_{j=q_1}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

$$AV_r(p) = 1 - \left[1 - \sum_{j=q_r}^{n/r} \binom{n/r}{j} p^j (1-p)^{n/r-j}\right]^r \ .$$

The inequality $AV_r(p) > AV_1(p)$ holds if and only if

$$\left[\sum_{j=0}^{q_r-1} \binom{n/r}{j} p^j (1-p)^{n/r-j}\right]^r < \sum_{j=0}^{q_1-1} \binom{n}{j} p^j (1-p)^{n-j} \tag{21}$$

Let $q = p/(1-p)$. Using the Vandermonde identity and following the derivation of the inequality (18) of Theorem 2, we find in this case

$$c_{q_{MAJ}}\, q^{q_{MAJ}} < \left[\binom{n}{q_r} - c_{q_r}\right] q^{q_r} .$$

Therefore

$$q^{r(r-1)m} < \left[\binom{n}{q_r} - c_{q_r}\right] / c_{q_{MAJ}} .$$

As before, we use only the dominant terms on the right. It suffices to set for large n

$$q^{r(r-1)m} < \binom{n/r}{q_r} / \binom{n}{q_{MAJ}} ,$$

or

$$q^{r(r-1)m} < \binom{n/r}{n/(k+r)} / \binom{n}{n/(k+1)} \tag{22}$$

Writing the right hand side of (22) in the form

$$\binom{(k+1)(k+r)m}{r(k+1)m} / \binom{r(k+1)(k+r)m}{r(k+r)m}$$

and using the approximation (9), it can be shown that for large n

$$q^{r(r-1)} < \frac{k^{rk(k+r)}(k+r)^{(k+1)(k+r)}}{r^{r(k+1)} k^{k(k+1)} (k+1)^{r(k+1)(k+r)}} ,$$

and therefore for large n and k

$$q < \frac{1}{r^{k/(r-1)}(k+1)^{(k+r)/(r-1)}} ,$$

which is the statement of the theorem. □