# 3 Quorum Systems

## 3.1 Introduction

**Motivation.** Quorum systems represent a fundamental abstraction for coordination among the nodes of a distributed system.

**Definition 1 (Quorum system).** Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a set of servers. A *quorum system* $\mathcal{Q} \subset 2^{\mathcal{P}}$ is a set of subsets of $\mathcal{P}$ such that every two subsets intersect. Each $Q \in \mathcal{Q}$ is called a *quorum*.

W.l.o.g., quorum systems considered here are minimal, i.e., for any $Q, Q' \in \mathcal{Q} : Q \not\subseteq Q'$.

**Algorithm 2 (Mutual exclusion).** Suppose the servers of $\mathcal{P}$ share a single resource that requires *mutually exclusive* access. Before entering the critical section, a server must obtain permission from all members of a quorum to proceed. The server does this by picking a quorum $Q$ and asking all its members for permission. When it leaves the critical section again, it informs the servers in $Q$ that it no longer accesses the resource.

Every keeps track if it has currently granted permission to enter the critical section. It gives permission only if it has not granted it, and marks it again as not granted when it is informed that the server has exited the critical section.

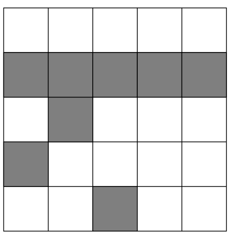
The intersection property of the quorum system ensures the integrity of the shared resource (safety). Note that servers may crash unless they are in the critical section. This algorithm may deadlock — liveness can be assured only by adding timeouts and additional messages [Mae85, AE89].

## 3.2 Example Quorum Systems

**Singleton.** $\mathcal{Q} = \big\{\{P_1\}\big\}$.

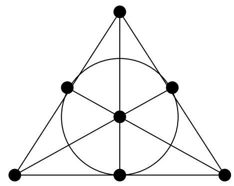
**Majority.** $\mathcal{P} = \{P_1, \ldots, P_n\}$ and $\mathcal{Q} = \big\{Q \subset \mathcal{P} \big| |Q| = \lceil \frac{n+1}{2} \rceil \big\}$; tolerates $t < \frac{n}{2}$ faulty servers. Generalization to *weighted majority* by assigning multiple "votes" to some servers.

**Grid.** Suppose $n = k \cdot k$ and arrange the servers in a square; a quorum is a full row and one element from each row below the full row.



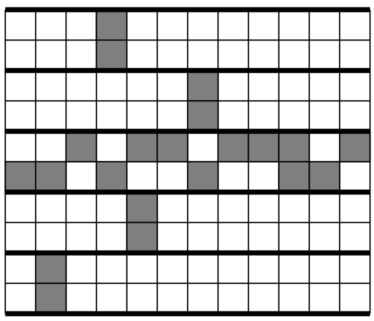
A grid quorum on $n = 25$ elements.

**Finite Projective Planes (FPP) [Mae85].** Suppose $n = q^2 + q + 1$ for a prime power $q$; then there is a finite projective plane on $n$ elements, which consists of a set of subsets from $\mathcal{P}$ such that every subset has exactly $q+1$ elements, every element is contained in exactly $q+1$ subsets, and every two subsets intersect in exactly one element.



The finite projective plane of order 2 ("Fano plane").

**B-Grid [NW98].** Suppose $n = dhr$ and arrange the elements in a grid with $d$ columns and $h \cdot r$ rows. Call every group of $r$ rows a *band* and call $r$ elements in a column restricted to a band a *mini-column*. A quorum consists of one mini-column in every band and one element from each mini-column of one band; thus, every quorum has $d + hr - 1$ elements.



The B-Grid quorum system over $n = 120$ elements with $d = 12$ columns, $h = 5$ bands, and $r = 2$ rows per band.

## 3.3 Measures on Quorum Systems

The load is a property of the quorum system; it is defined as the probability that the *busiest* server is in use under an *optimal* strategy of accessing the servers.

**Definition 3 (Load).** An *access strategy* $W$ is a random variable on a quorum system $\mathcal{Q}$, i.e., $\sum_{Q \in \mathcal{Q}} P_W(Q) = 1$. The *load induced by $W$ on a server $P_i$* is

$$\ell_W(i) \;=\; \sum_{Q \in \mathcal{Q}: P_i \in Q} P_W(Q).$$

The *load induced by $W$ on a quorum $\mathcal{Q}$* is the maximal load induced by $W$ on any server in $\mathcal{Q}$, i.e.,

$$L_W(\mathcal{Q}) \;=\; \max_{P_i \in \mathcal{P}} \ell_W(i).$$

The *[system] load of $\mathcal{Q}$* is

$$L(\mathcal{Q}) \;=\; \min_W L_W(\mathcal{Q}).$$

Let $c(\mathcal{Q})$ denote the size of the smallest quorum of a quorum system $\mathcal{Q}$.

**Theorem 4 ([NW98]).** *$L(\mathcal{Q}) \geq \max\{\frac{1}{c(\mathcal{Q})}, \frac{c(\mathcal{Q})}{n}\}$. Consequently, $L(\mathcal{Q}) \geq \frac{1}{\sqrt{n}}$.*

Resilience is a worst-case measure for the fault-tolerance of a quorum system, defined as the maximum number of faulty servers that the quorum system can tolerate.

**Definition 5 (Resilience).** The *resilience* $R(\mathcal{Q})$ of a quorum system is the largest $f$ such that for all sets $F \subset \mathcal{P}$ of cardinality $f$, there is at least one quorum $Q \in \mathcal{Q}$ which has no common element with $F$.

Clearly, the resilience is at most $c(\mathcal{Q}) - 1$.
An average-case measure for fault-tolerance is the failure probability. Assume that every server $P_i$ fails independently with probability $p$; let $FAIL(i)$ denote the event that $P_i$ fails.

**Definition 6 (Failure probability).** The *failure probability* of a quorum system $\mathcal{Q}$ is the probability that at least one server of every quorum fails, i.e.,

$$F_p(\mathcal{Q}) \;=\; \Pr\big[\forall Q \in \mathcal{Q} : \exists P_i \in Q \;\text{ such that } FAIL(i)\big].$$

**Definition 7 ($s$-uniform).** A quorum system $\mathcal{Q}$ is *$s$-uniform* if every quorum in $\mathcal{Q}$ has exactly $s$ elements.

**Definition 8 (Balanced access strategy).** An access strategy $W$ for a quorum system $\mathcal{Q}$ is *balanced* if it satisfies $\ell_W(i) = L$ for all $P_i \in \mathcal{P}$.

**Lemma 9.** *An $s$-uniform quorum system $\mathcal{Q}$ with a balanced access strategy has load $L(\mathcal{Q}) = L = \frac{s}{n}$. Moreover, this load is optimal.*

*Proof.* [NW98, Proposition 4.8]. □

**Lemma 10.** *The B-Grid quorum system has load $\frac{d+hr-1}{dhr}$, resilience $\min\{d-1, hr-1\}$, and failure probability at most $(dp^r)^h + h(1-(1-p)^r)^d$. For $d = \sqrt{n}$, $r = \lfloor \ln d \rfloor$, and $0 \leq p \leq \frac{1}{3}$, we have $L(\text{B-Grid}) = O(\frac{1}{\sqrt{n}})$, $R(\text{B-Grid}) = O(\sqrt{n})$, and $F_p(\text{B-Grid}) = O\big(e^{-\frac{n^{1/4}}{2}}\big)$.*

*Proof.* Load follows from Lemma 9 since B-Grid is $(d + hr - 1)$-uniform. Define $\mathcal{E}_1$ to be the event that *in every band, all elements of some mini-column fail*, and $\mathcal{E}_2$ the event that *in some band, at least one element of every mini-column fails*. Clearly, the system fails when $\mathcal{E}_1 \vee \mathcal{E}_2$ and thus

$$F_p(\text{B-Grid}) \;\leq\; \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2] \;=\; (dp^r)^h + h(1-(1-p)^r)^d.$$

The last expression is bounded by $O\big(e^{-h} + e^{-\frac{\sqrt{d}}{2}}\big)$, which is bounded by $O\big(e^{-\frac{n^{1/4}}{2}}\big)$ for sufficiently large $n$. □

**Comparison.**

| $\mathcal{Q}$ | $L(\mathcal{Q})$ | $R(\mathcal{Q})$ | $F_p(\mathcal{Q})$ |
|---|---|---|---|
| Singleton | $1$ | $0$ | $p$ |
| Majority | $\frac{1}{2}$ | $\lfloor \frac{n-1}{2} \rfloor$ | $e^{-\Omega(n)}$ |
| Grid | $O(\frac{1}{\sqrt{n}})$ | $\sqrt{n} - 1$ | $\approx 1^*$ |
| FPP | $O(\frac{1}{\sqrt{n}})$ | $q$ | $\approx 1^*$ |
| B-Grid$^+$ | $O(\frac{1}{\sqrt{n}})$ | $O(\sqrt{n})$ | $O\big(e^{-\frac{n^{1/4}}{2}}\big)$ |

$^{*}$ for large $n$.

$^{+}$ for $d = \sqrt{n}$, $r = \lfloor \ln d \rfloor$, and $0 \leq p \leq \frac{1}{3}$.

Only the B-Grid quorum system achieves optimal and close-to-optimal values of all three measures.

## References


[AE89] D. Agrawal and A. El Abbadi, *An efficient solution to the distributed mutual exclusion problem*, Proc. 8th ACM Symposium on Principles of Distributed Computing (PODC), 1989, pp. 193–200.

[Mae85] M. Maekawa, *A $\sqrt{N}$ algorithm for mutual exclusion in distributed systems*, ACM Transactions on Computer Systems **3** (1985), no. 2, 145–159.

[NW98] M. Naor and A. Wool, *The load, capacity and availability of quorum systems*, SIAM Journal on Computing **27** (1998), no. 2, 423–447.