Adaptive Page Allocation: Cost and Replication Analysis

Maksymilian Neumann

June 5, 2025

Abstract

We analyze the performance of the COUNT replication algorithm under varying write probabilities p and replication thresholds D. We measure two primary metrics: the average cost per request and the average maximum number of replicas. Results are presented via line plots and heatmaps to illustrate trends and trade-offs.

1 Introduction

In distributed shared memory systems, dynamic replication can reduce read and write costs by adaptively placing copies of data. The *COUNT* algorithm tracks writes per node and replicates once a threshold D is exceeded. We evaluate how write intensity p and threshold D affect:

1. the average cost per request,
2. the maximum number of replicas observed during execution.

2 Methodology

We simulate a 64-node fully connected network. For each (D, p) pair, we run 10,000 independent simulations of 65,536 requests, randomly drawn from a uniform distribution over nodes. Each request is:

- a **write** with probability p,
- a **read** with probability $1 - p$.

Costs:

- **Read:** Cost is 0 if a local replica exists; 1 otherwise.
- **Write:** Cost is number of replicas if the page is not replicated, or one less if it is.
- **Replication:** When local counter reaches D, a copy is placed at cost D. Counter resets.
- **Decay/Eviction:** Counters decay on writes to other pages; replicas may be evicted if counters drop to 0.

The experiment is implemented in Rust using multithreading (via Rayon) and randomness (via `rand`). Data are exported to CSV and visualized via heatmaps and line plots.

3 Results

3.1 Average Cost per Request

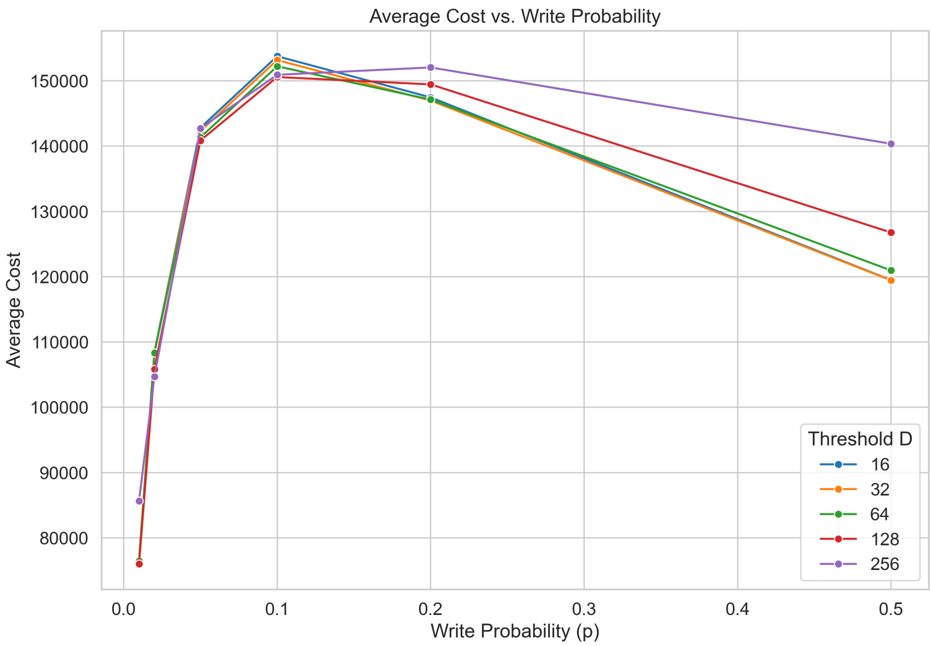

Figure 1: Average request cost vs. write probability p for various thresholds D.

3.2 Average Maximum Number of Replicas

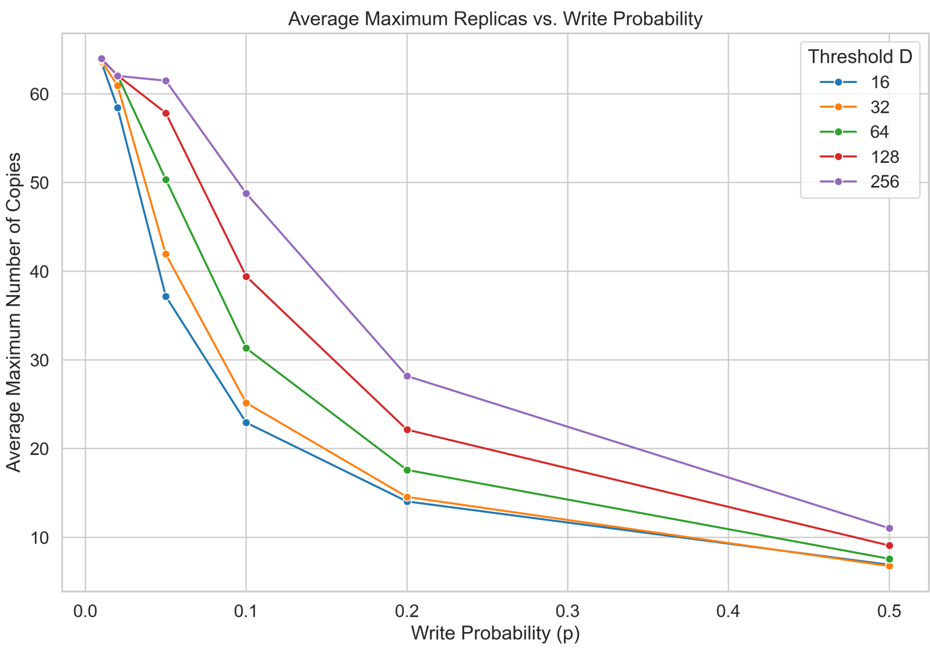

Figure 2: Average maximum number of replicas vs. write probability p.

3.3 Heatmap of Average Cost

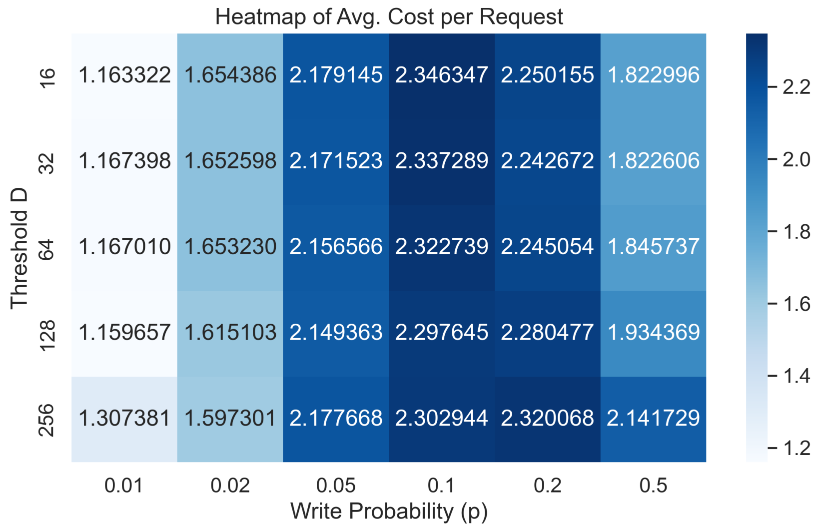

Figure 3: Heatmap showing average cost per request across thresholds D and write probabilities p.

3.4 Heatmap of Average Max Copies

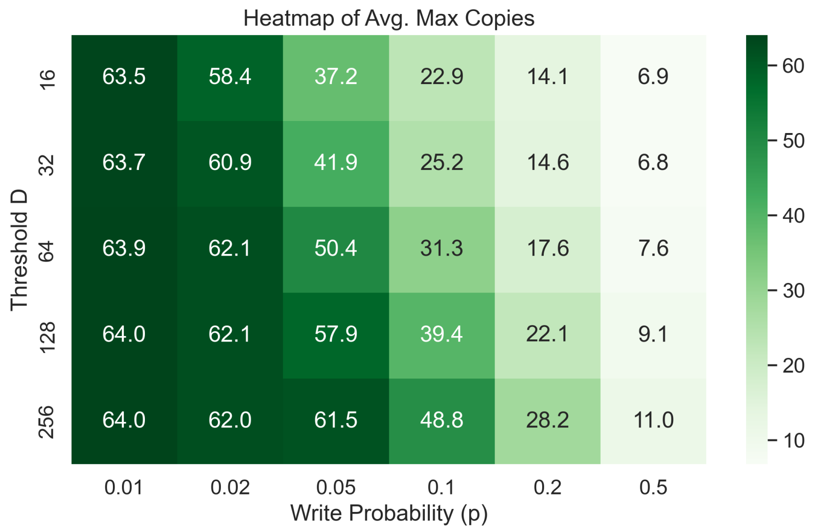

Figure 4: Heatmap showing average maximum replication degree as function of D and p.

4 Discussion

- **Cost dynamics:** As observed in Figure 1, all thresholds exhibit a peak in cost at moderate p values ($0.1 \leq p \leq 0.2$), after which the cost decreases. This is due to a balance shift: at low p, replication reduces read costs, while at high p, fewer nodes are replicated due to fast eviction.
- **Unexpected replica saturation:** Unlike earlier simulations, high values of D (e.g., 256) now result in large replication degrees for $p \leq 0.1$, sometimes reaching nearly 64 copies.

Figure 4 confirms that under low write intensity, replication accumulates regardless of threshold due to slow decay.

- **Non-monotonic cost behavior:** Heatmap in Figure 3 shows that average request cost increases for $p \in [0.02, 0.2]$, especially when $D \leq 64$, due to elevated write costs from large replica sets.
- **Choosing D:** For minimal cost, thresholds in the 32–128 range offer reasonable trade-offs. Very low thresholds ($D = 16$) replicate too aggressively. Surprisingly, $D = 256$ causes high costs at moderate p, due to wide replication before decay kicks in.
- **Implications:** System designers should match D to expected access patterns. If reads dominate and decay is infrequent, small D is risky. Conversely, high D can bloat replication during low-write phases.