Scheduling under redundancy

E. Anton[1], R. Righter[2] and I.M. Verloop[3,4]

[1]Eindhoven University of Technology, Eindhoven, The Netherlands
[2]University of California at Berkeley, Berkeley, USA
[3]CNRS - IRIT, Toulouse, France
[4]Université de Toulouse - INP, Toulouse, France

In the present extended abstract we investigate the impact that the scheduling policy has on the performance of redundancy systems when the usual exponentially distributed i.i.d. copies assumption is relaxed. In particular, we investigate the performance, in terms of the total number of jobs in the system, not only for redundancy-oblivious policies, such as FCFS (First-Come-First-Serve) and ROS (Random-Order-of-Service), but also for redundancy-aware policies of the form $\Pi_1 - \Pi_2$, where Π_1 discriminates among job classes and Π_2 discriminates among jobs of the same class. Examples of first-level policies are LRF (Least-Redundant-First) and MRF (Most-Redundant-First), where under LRF, respectively MRF, within a server jobs with fewer copies, respectively more copies, have priority over jobs with more copies, respectively fewer copies. Second-level policies could be FCFS or ROS.

Under the cancel-on-complete (*c.o.c.*) redundancy model, an arriving job dispatches multiple copies to all compatible servers, and departs when the first copy completes service. Redundancy aims to exploit the variability of the queue lengths and server capacities, potentially reducing the response time.

The stability of redundancy models, which is the first performance measure, has been studied in recent work and is summarized in Anton et al. [2]. Under FCFS and when jobs have independent and identically distributed (i.i.d.) copies, the stability region is not reduced due to adding redundant copies. The latter also holds for the so-called *redundancy-d* model under the ROS and PS policies. When copies are instead identical, the stability condition strongly depends on the scheduling policy implemented in the servers. In particular, for exponential service times, the stability region of the redundancy-d model is not reduced when the scheduling policy is ROS, but it is dramatically reduced when the scheduling policy is either FCFS or PS.

The impact of the redundancy policy on the number of jobs in the system was first studied in [5, 6] for the FCFS scheduling policy, where each job can dispatch i.i.d. copies to any server in the system. Assuming NWU service time distributions, Koole and Righter [6] show that full replication stochastically minimizes the number of jobs in the system at any time. In contrast, for NBU service time distributions, Kim et al. [5] show that no replication is optimal.

In [3, 4], redundancy-aware policies are introduced. These papers investigate the impact that the implemented scheduling policy has on the performance for nested redundancy models with exponential service times and i.i.d. copies. In [3], the authors consider the W-model and prove that implementing FCFS in the servers is highly effective in reducing the mean response time in the system, even though LRF is optimal. However, LRF fails to be fair to non-redundant jobs. In Gardner et al. [4], the authors consider general nested systems and show that for LRF even if scheduling more redundant jobs is better, the maximum gains come from adding only a small proportion of redundant jobs.

In this extended abstract, we assume that jobs have general service times and consider both the cases where copies are independent and when they identical . Assuming that jobs have independent and identically distributed (i.i.d.) copies, we show the following: (i) When jobs have exponential service times, LRF policies outperform any other policy. (ii) When service times are New-Worse-than-Used, MRF-FCFS outperforms LRF-FCFS as the variability of the service time grows infinitely large. (iii) When service times are New-Better-than-Used, LRF-ROS (resp. MRF-ROS) outperforms LRF-FCFS (resp. MRF-FCFS) in a two-server system. Statement (iii) also holds when job sizes follow a general distribution and have identical copies (all the copies of a job have the same size). Moreover, we show via simulation that, for a large class of redundancy systems, redundancy-aware policies can considerably improve the mean response time compared to redundancy-oblivious policies. We refer to the technical report [1] for further details.

1. MODEL DESCRIPTION

We consider a K parallel server system with heterogeneous capacities μ_s, for $s \in S$, where $S = \{1, \ldots, K\}$ is the set of all servers. Jobs arrive to the system according to a Poisson process of rate λ. Each job is independently labelled with a class c that represents the subset of servers to which it sends a copy: i.e., $c = \{s_1, \ldots, s_n\}$, where $n \leq K$, $s_1, \ldots, s_n \in S$ and $s_i \neq s_l$, for all $i \neq l$. We denote by $\mathcal{C}$ the set of all classes in the system. An arriving job is with probability p_c of class c, with $\sum_{c\in\mathcal{C}} p_c = 1$. We assume that the redundancy topology is nested, that is, the set of classes $\mathcal{C}$ satisfies the following: for all job classes $c, c' \in \mathcal{C}$, either *i)* $c \subseteq c'$ or *ii)* $c' \subseteq c$ or *iii)* $c \cap c' = \emptyset$. The W-model is a particular nested system; this is a $K = 2$ server system with job classes $\mathcal{C} = \{\{1\}, \{2\}, \{1, 2\}\}$.

We denote by π a generic scheduling policy. We assume that the policy π has no information on the actual size of the copies, that is, it is non-anticipating. We introduce two-level redundancy-aware scheduling policies $\pi = \Pi_1$-Π_2:

- The first-level policy Π_1 determines the preemptive priority among job classes.
- The second-level policy Π_2 determines the scheduling policy of jobs within the same class. This policy is assumed to be size-unaware and non-preemptive within the class. That is, once a job of a given class is started at a server, no other job of the same class can be served at that server until the given job has completed (at some server).

Examples of first-level policies Π_1 are Least-Redundant-First (LRF) and Most-Redundant-First (MRF). Examples of second-level policies Π_2 are FCFS and ROS. The latter are also examples of single-level redundancy-oblivious policies.

For a given scheduling policy π, we denote by $N^{\pi}(t)$ the total number of jobs present in the system at time t.

We aim to compare the performance of the system under different scheduling policies. We have the following stochastic ordering definition.

Definition 1. *For two nonnegative continuous random variables X and Y, with respective cumulative distributions F and G, and $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(x) = 1 - G(x)$, we say that $X \geq_{st} Y$, that is, X is stochastically larger than Y, if $\bar{F}(x) \geq \bar{G}(x)$ for all $x \geq 0$.*

We let X denote the service time distribution of a job when it is served at capacity 1. Special focus will be given to exponential service times, as well as the following two classes of service time distributions: New-Worse-than-Used (NWU) and New-Better-than-Used (NBU), defined as follows. Let $X_t = [X - t | X > t]$ be the remaining processing time of a job that has completed t time units of service.

Definition 2. *We say that X is New Worse than Used (NWU), resp. New Better than Used (NBU), if the remaining processing time of a task that has received some processing is stochastically* larger*, resp.* smaller*, than the processing time of a task that has received no processing, i.e., $X_0 \leq_{st} X_t$ for all t, resp., $X_t \leq_{st} X_0$ for all t.*

2. STOCHASTIC COMPARISON RESULTS

In this section, we analyze how the scheduling policy affects the performance of the system.

2.1 Exponential service times and i.i.d. copies

We first assume that service times are exponentially distributed with i.i.d. copies. We can generalize the result in [3] for LRF-FCFS, because of the memoryless property, which implies that the number of jobs is insensitive to the implemented second-level policy Π_2.

Proposition 1. *Consider a redundancy system with a nested topology and heterogeneous server capacities where jobs have exponentially distributed i.i.d. copies. Then,*

$$\{N^{LRF-\Pi_2}(t)\}_{t\geq 0} \leq_{st} \{N^{\pi}(t)\}_{t\geq 0},$$

for any Π_2 and any π.

We note that for non-nested topologies, an optimal policy is expected to be more complex because an optimal choice of which class to serve will depend on the number of jobs in each class.

2.2 NWU and i.i.d. copies

When jobs have i.i.d. copies and NWU service time distributions, the service time of a copy that is already in service is stochastically larger than that of an i.i.d. copy that has not received service yet. Hence, this suggests that whenever a server becomes available to a class, it will be better to serve a copy of a job that has already a copy elsewhere in service, because the new copy has a good chance of completing sooner. That is exactly what policy Π_2 =FCFS does. In the result below we show that, for any given first-level policy Π_1, FCFS is indeed optimal. We note that in [6] this result was proved for the redundancy system with only one class of jobs.

Proposition 2. *Consider a redundancy system with a nested topology, heterogeneous server capacities, NWU service times and i.i.d. copies. Then,*

$$\{N^{\Pi_1-FCFS}(t)\}_{t\geq 0} \leq_{st} \{N^{\Pi_1-\Pi_2}(t)\}_{t\geq 0},$$

for all $t \geq 0$ and any first second-level policy Π_1.

As an illustration, in Figure 1 we simulate the W-model with $\lambda = 1.3$, homogeneous capacities $\vec{\mu} = (1, 1)$ and $p_c = 1/3$ for all $c \in \mathcal{C}$. We assume that the service time distribution X is a mixture of Y/q with probability q and 0 otherwise, where $Y \sim F$ is NWU. In Figure 1 we chose Y having an exponential distribution. The coefficient of variation of X equals $C^2 = \frac{\mathbb{E}(Y^2)}{q\mathbb{E}(Y)^2} - 1$ and increases without bound when $q \to 0$. We note that in the special case where Y is exponentially distributed, the coefficient of variation equals $C^2 = 2/q - 1$. Consistent with Proposition 2, we observe that Π_1-FCFS (solid line) outperforms Π_1-ROS (dashed line) for both Π_1=LRF ($\times$) and Π_2=MRF ($\circ$). This observation also holds for single-level redundancy-oblivious policies, i.e., FCFS outperforms ROS, however, we did not obtain a proof for this.

In Figure 1 we also observe that as q approaches 1, LRF-FCFS outperforms the other policies. In fact, when $q = 1$ and Y is exponentially distributed, it was shown in Proposition 1 that LRF-FCFS minimizes the number of jobs. When q approaches 0, that is $C^2 \to \infty$, we observe that MRF-FCFS outperforms all other scheduling policies.

This example shows that under NWU service times there is not a unique first-level policy that minimizes the total number of jobs in the system for non-exponential service times. MRF often performs well, because this policy maximizes the number of copies of the same job in service. On the other hand, LRF tend to minimize the time that servers are idle. Hence, there is a trade-off, and which policy is optimal will strongly depend on the coefficient of variation of the service time distribution, which impacts how beneficial it is to serve copies of the same job (the more variable services are, the more profitable to serve copies of the same job). The proposition below supports our observation for a K server system where each server has dedicated traffic and there is one flexible class of jobs that sends copies to all servers.

Proposition 3. *Consider K heterogeneous servers with capacities μ_s where each server has a dedicated job class and there is an additional job class that sends copies to all the servers. That is, $\mathcal{C} = \{\{s\}_{s\in S}, S\}$. We assume that copies are i.i.d. and that the service time distribution X is a mix-*

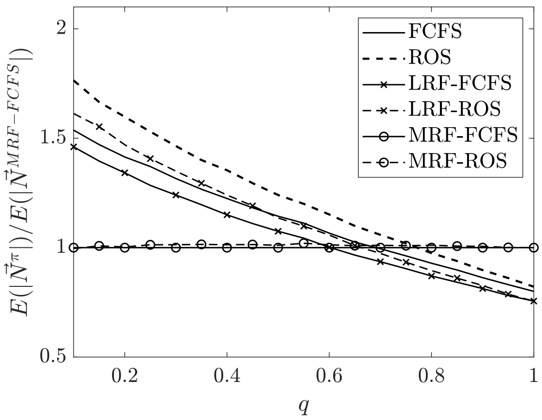

Figure 1: The mean number of jobs for the W-model with i.id. copies when $\vec{\mu} = (1,1)$, $\lambda = 1.3$ and $p_c = 1/3$ for all $c \in \mathcal{C}$ with X a mixture of exponential service times with respect to q (NWU).

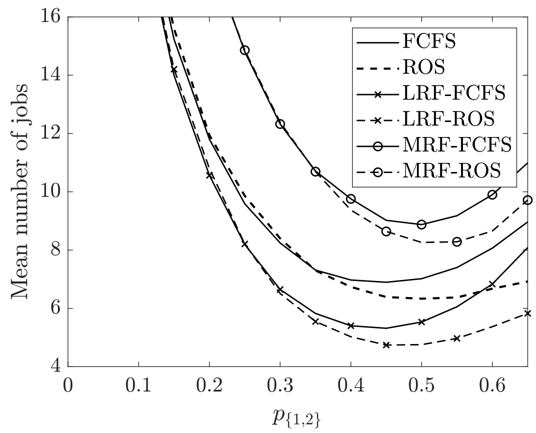

Figure 2: The mean number of jobs for the W-model with identical copies, when $\lambda = 1.3$, $\vec{\mu} = (2,1)$, $p_{\{1\}} = 0.35$, $p_{\{2\}} = 1 - p_{\{1\}} - p_{\{1,2\}}$. Mixture of exponential service times with $q = 0.1$ (NWU).

ture of Y/q with probability q and 0 otherwise, where $Y \sim F$ is NWU. Then,

$$q\mathbb{E}(N^{\text{MRF-FCFS}}) < q\mathbb{E}(N^{\text{LRF-FCFS}}) + o(1), \text{ as } q \to 0.$$

Moreover, in Figure 1 we observe that the gap between Π_1-ROS and ROS, and the gap between Π_1-FCFS and FCFS, increases as the variability of the service time increases. The redundancy-oblivious policies can be more than a factor 1.5 worse than the MRF redundancy-aware version.

2.3 NBU and i.i.d. copies or general service times and identical copies

In this section we assume that jobs have either i.i.d. copies and NBU service times, or identical copies and follow a general distribution.

We note that for these particular systems, having several copies of the same job in service implies that the capacity of one of the servers is unnecessarily dedicated to this job. The following proposition gives a partial characterization for an optimal second-level policy for a two-server system.

Proposition 4. *Consider a W-model with heterogeneous servers, and either NBU service times and i.i.d. copies or general service times and identical copies. The first-level policy is hence either LRF or MRF. Whenever the first-level priority policy serves class $\{1,2\}$ and the second-level policy decides not to idle, it is stochastically better to schedule at the second level according to ROS than to FCFS.*

Note that the above proposition does not give conditions under which non-idling or idling is optimal.

In Figure 2 we compare the different policies (without idling) and observe that for a given first-level policy, ROS outperforms FCFS. We also observe that LRF-FCFS outperforms FCFS, and LRF-ROS outperforms ROS. Indeed, the redundancy-aware version can be a factor 1.25 better than the redundancy-oblivious policy.

Numerically we observe that LRF outperforms MRF, see Figure 2. In the case of identical copies and general service times, or deterministic service times, we can indeed prove this under certain conditions.

Proposition 5. *Consider K heterogeneous servers with capacities μ_s where each server has a dedicated job class and there is an additional job class that sends copies to all the servers. That is, $\mathcal{C} = \{\{s\}_{s\in S}, S\}$. We assume that $\mu_1 = \max_{s\in S}\{\mu_s\}$. Jobs have general service times and identical copies, or deterministic service times. Assume that $p_S \geq p_{\{1\}}$ or that the arrival rate λ is small enough. Then it holds that*

$$\mathbb{E}(N^{LRF-FCFS}) \leq \mathbb{E}(N^{MRF-FCFS}).$$

3. REFERENCES

[1] E. Anton, R. Righter, and I. M. Verloop. Efficient scheduling in redundancy systems with general service times. *Under submission*, 2022.

[2] E. Anton, U. Ayesta, M. Jonckheere, and I. M. Verloop. A survey of stability results for redundancy models. In *Alexey B. Piunovskiy and Yi Zhang (eds.), Modern Trends in Controlled Stochastic Processes: Theory and Applications, V.III.* Springer US, 2021.

[3] Kristen Gardner, Mor Harchol-Balter, Esa Hyytia, and Rhonda Righter. Scheduling for efficiency and fairness in systems with redundancy. *Performance Evaluation*, 116: 1–25, 2017.

[4] Kristen Gardner, Esa Hyytiä, and Rhonda Righter. A little redundancy goes a long way: Convexity in redundancy systems. *Performance Evaluation*, 131(4): 22–42, 2019.

[5] Yusik Kim, Rhonda Righter, and Ronald Wolff. Job replication on multiserver systems. *Advances in Applied Probability*, 41: 546–575, 2009.

[6] Ger Koole and Rhonda Righter. Resource allocation in grid computing. *Journal of Scheduling*, 11: 163–173, 2008.