Locally Balanced Allocations under Strong Byzantine Influence

Costas Busch[1] [0000−0002−4381−4333], Paweł Garncarek[2] [0000−0002−6855−0530], and Dariusz R. Kowalski[1] [0000−0002−1316−7788]

[1] School of Computer and Cyber Sciences, Augusta University, USA
{kbusch,dkowalski}@augusta.edu
[2] Institute of Computer Science, University of Wrocław
pawel.garncarek2@uwr.edu.pl

Abstract. The Power of Two Choices (PoTC) is a commonly used technique to balance the incoming load (balls) into available resources (bins) – for each coming ball, two bins are selected uniformly at random and the one with smaller number of balls is chosen as the location of the current ball. We study a generalization of PoTC to a fault-prone setting – faulty bin(s) could present malicious information to enforce allocation decision on any of the two bins. Given m balls and n bins, such that no more than f of the bins are faulty, we show that the maximum loaded honest bin receives a surplus of a logarithmic number of balls with respect to f. Our result generalizes the classic bounds of the Power of Two Choices in the presence of a strong Byzantine adversary. Our solution and methods of analysis can help to efficiently implement and analyze resilient online local decisions made by processes when solving fundamental problems that depend on load balancing under the presence of Byzantine failures.

Keywords: Scheduling and resource allocation · Power of Two Choices · Fault tolerance · Strong Byzantine adversary · Probabilistic analysis.

1 Introduction

We study a generalization of the classical Power of Two Choices (PoTC) online principle (in which the ball chooses the bin with a smaller number of balls) to a fault-prone setting – a faulty bin could present the ball with malicious information to enforce the decision on any of the two bins. This generalization could allow using the PoTC paradigm in the design and analysis of distributed protocols, for example, population protocols, multiprocessor job scheduling, and network routing, under the presence of Byzantine failures.

In the traditional balls-in-bins problem, there are m balls and n bins and each ball is thrown into a bin picked uniformly at random [20,25]. It is well known that for $m = n$ each bin gets at most $\log n/\log\log n$ balls with high probability (i.e., with probability at least $1-o(1)$) [17]. This improves to at most $\log\log n + \Theta(1)$ balls per bin, if we give two random choices to each ball and the ball goes into the bin that has the smaller number of balls [5]. The result can be generalized for d random choices per ball to at most $(1+o(1))\log\log n/\log d+\Theta(m/n)$ balls per bin [5,11].

The balls-in-bins problem has important applications in fundamental decision-making problems [27]. It can be used to solve task load balancing, hash collision, and routing problems. In the task load balancing problem, each ball corresponds to a computing task, where each task has a fixed weight, and the bins are the servers that will execute the tasks. Using the Power of Two Choices,[3] each task picks two random servers and is allocated to the least loaded server. Hence, with n tasks and n servers, each server will be imbalanced by at most an additive $O(\log\log n)$ term. One of the advantages of PoTC is that it is local, i.e., decisions are made by looking into only 2 bins rather than all of them, which is important in distributed settings. A similar approach with PoTC can be used to improve collisions in hash functions [27]. With two perfectly random hash functions giving two distinct table entries, a key is placed to the entry with the smallest chain of keys. The load balancing attribute can also be used in network routing where packets pick two randomly chosen paths and the packet is sent along the path that currently has the lowest congestion [15]. Other applications include load-balancing in virtual machines in fog computing [10], queuing theory analysis [12], and distributed voting [16].

Many of the aforementioned applications for the PoTC problem take place in a distributed computing environment that is prone to failures [1,24]. The failures may take different forms, for example, the computing nodes may crash and communication links may be disrupted. Even worse, nodes and communication links may be compromised and misbehave due to malicious attacks or other reasons. Since typical distributed systems need to continuously operate even under failures, it is important to understand what is the system's resiliency to failures and the impact on the performance of the underlying distributed task.

In this work, we model failures as being generated by a Byzantine adversary that aims to disrupt normal operations and cause the system to misbehave. Byzantine failures model a large range of system failures which include crash failures and malicious attacks. Byzantine adversaries have been considered for classic distributed consensus [9,22]. Here, we examine the impact of Byzantine faults on PoTC.

The Byzantine adversary controls $f < n$ of the bins which are faulty and they may report wrong information about the number of balls in them. The remaining $n - f$ bins are honest and not controlled by the adversary. The Byzantine adversary does not control the scheduler, that is, the scheduler keeps making random picks of two bins out of n for each of the balls. If the scheduler picks two honest bins, the ball will fall into the bin with the least load. However, if the scheduler picks one honest bin together with a faulty bin, then the faulty bin may report a higher number of balls than it actually has, forcing the ball to fall into the honest bin. Therefore, a Byzantine adversary can create an imbalance in the allocation of balls into the bins, by directing more balls to the honest bins.

[3] "Power of Two Choices" can denote a principle as well as a greedy algorithm following said principle.

	Number of balls in a bin	m	Additional conditions	Prob.
Th. 1	$(1+\delta)\frac{2m}{n}$	any m	$\frac{2m\delta^2}{3n} > c\log n$ $\delta \in (0,1)$ any f	$1-n^{1-c}$
Cor. 1	$\frac{3m}{n}$	$m > 9n\log n$	any f	$1-n^{-1/2}$
Th. 2	$\phi + \log\log\frac{n^2}{2emf} +$ $+\log_{\frac{n^2}{2emf}} \frac{2f}{\sqrt{f^2+8\frac{n^2}{m}\log n - f}} + 1$	$m = O(n\log n)$	$1 \leq f < \frac{n^2}{2em}$	$1-o(1)$
Cor. 3	$O\left(\log\log\frac{n^2}{mf} + \log_{\frac{n^2}{2emf}} \frac{f}{2e\log n}\right)$	$m = O(n)$	$1 \leq f < \frac{n^2}{2em}$	$1-o(1)$
Cor. 4	$O(\frac{m}{n} + \log\log\frac{n^2}{mf} +$ $+\log_{\frac{n^2}{2emf}} \frac{f}{2e\log n})$	$m \geq en$ $m = O(n\log n)$	$1 \leq f < \frac{n^2}{2em}$	$1-o(1)$
Cor. 5	$O\left(\frac{m}{n} + \log\log\frac{n}{f} + \log f\right)$	$m \geq n$	$1 \leq f <$ $\max\{\frac{n}{12e\log n}, \frac{n^2}{2em}\}$	$1-o(1)$

Table 1. Technical contributions. Parameter ϕ can be equal to $\frac{4em^2}{n^2}$ (see Corollary 2) or cem/n for any $c > 2e$ (see Corollary 5 and Lemma 7).

We show that PoTC has an inherent tolerance to failures. Given f faulty bins, for the case where $f < n^2/(2em)$ and $m = O(n\log n)$, we show that the maximum loaded honest bin receives only a logarithmic number of balls with respect to f. We also bound the maximum load on an honest bin for the general case. Our results imply that the system can gracefully tolerate failures for certain values of f without significantly affecting the maximum load on any bin. Hence, the affected applications related to load-balancing and collision avoidance have sustained good properties with only an additive logarithmic term on their performance impact due to failures.

Our main technical contributions are summarized in Table 1. All the results occur with high probability. The main results are provided by Theorem 1 and Theorem 2. All the corollaries are derived from these two theorems. Theorem 1 applies to a wide range of number of balls m. A more precise result is shown in Theorem 2, however it applies only to $m = O(n\log n)$ and depends on a parameter ϕ that can be used to provide refined results. Depending on the choice of the problem parameters, and replacing parameter ϕ accordingly, one can derive from Theorem 2 results that can be compared to the results for non-faulty PoTC, e.g., $m/n + O(\log\log n)$ in [11]. A precise description of these results, as well as some additional results and a discussion of why simpler solutions do not work, can be found in Section 2.

1.1 Related work

There are several variants of the PoTC problem. The extension to d choices provides an improvement to the basic $\log\log n$ maximum load bound [5,11]. It is shown that the improvement is inversely proportional to d if there is asymmetry between the d choices [28]. Typically it is assumed that all balls have the same weight, but having different ball weights has also been considered [26], since it represents the load-balancing problem of non-uniform tasks. Variants allowing ball deletions together with insertions have also been studied [7].

Recently, fault-tolerance of PoTC has been studied as well. For example, [23] focuses on load balancing applications and allows incorrect results of comparisons when the compared values differ by less than some parameter.

The classic PoTC problem has been generalized to graphical allocations [6,21], where the bins are nodes in a graph and for each ball, a uniformly random edge is selected. The ball is allocated to the less loaded of the two bins associated with the edge. The classic problem is a special case when the graph is the clique K_n. It is shown that in a n^{ϵ}-regular graph (with n nodes) the maximum load is $\log\log n + O(1/\epsilon) + O(1)$. Graphical allocations have also been studied from the perspective of assigning the ball to the lowest loaded bin with probability β and to the other bin with probability $1-\beta$, also called $(1+\beta)$-choice process [26]. Notably, $(1+\beta)$-choice process can be used to model faults caused by an unreliable load reporting mechanism. They show that the gap between the most and least loaded bins is $\Theta(\log n/\beta)$ irrespective of m (w.h.p.). Faults can be modeled with $(1+\beta)$-choice process for the case where an adversary always gives the ball to the honest node when also a faulty node is picked in the pair. However, a Byzantine adversary can behave in a way that cannot be mimicked by the random $(1+\beta)$-choice process, e.g. deterministically choosing balls that will be put in faulty bins. Additionally, our results are better for some parameters m, n, f than the results for $(1+\beta)$-choice process (see the note after Corollary 5).

PoTC naturally relates to population protocols [3,8], where there are n nodes (similar to the bins) such that a random scheduler each time picks two nodes to participate in an *exchange* (similar to picking two random bins to throw the ball). Population protocols solve basic distributed problems such as leader election and majority consensus using a small set of states per node. A useful primitive in population protocols is the *phase clock* [2],which is a local counter at each node that increments at each exchange modulo a fixed phase duration, such that each phase implements a step of an algorithm. This is similar to the behavior of PoTC as well. There is a relation with faulty population protocols, where some nodes may be malicious or act erratically [4,13,18]. In such settings, a fault-tolerant adaptation of phase clock needs to be developed. Such a phase clock requires bounds on the gap between the most loaded and least loaded bin. Our analysis of the most loaded bin of faulty PoTC is a step towards implementing a fault-tolerant phase clock for fault-resilient population protocols; however, our analysis needs to be adapted to bound the gap for phase clocks.

Outline of the paper: We continue with the problem specification, description of challenges and our extended results in Section 2. The proof of Theorem 2 is given in Section 3. The case $m \geq en$ is analyzed in Section 4, while the case of large m is analyzed in Section 5. Finally, we conclude with a discussion and open problems in Section 6.

Algorithm 1: Balls in bins with Power of Two Choices against strong Byzantine adversary

Input : n bins and a sequence of m balls $r_1, r_2, \dots, r_m$; Byzantine adversary controls f bins
Output: Distribution of the m balls to the bins

Let v_i denote the actual number of balls in bin i;
Initially, $v_i \leftarrow 0$, for all $i \in [n]$;
Let F, with $|F| = f$, be the set of bins controlled by the adversary;
For each $i \in F$, the value v_i is reported by the adversary;
For each $i \in [n] \setminus F$, the reported value v_i is the true number of balls in bin i;
for $k = 1$ *to* m **do**
    Pick uniformly at random a pair $(i, j) \in [n]^2$, $i \neq j$;
    `// Adversary knows the true values of` v_i `and` v_j
    `// If` $i \in F$ `or` $j \in F$ `then the adversary may report wrong` v_i `or` v_j`, resp.`
    Let x be the bin such that $x \in \{i, j\}$ and $v_x = \min(v_i, v_j)$;
    Place ball r_k into bin x;
    $v_x \leftarrow v_x + 1$;

2 Problem Specification and Our Results

2.1 Problem description

Consider n bins and m balls. The balls arrive one-by-one in an arbitrary linear order. The goal of an online algorithm is to throw each ball to some bin in such a way that the size of the largest bin (i.e., with the largest number of balls thrown to it) is minimized. f of the n bins are *faulty*, while all other bins are called *honest*.

In this work, we study a fault-tolerant version of an algorithm commonly known as the Power of Two Choices, called *Byzantine-Tolerant Power of Two Choices* or *BT_PoTC* for short. It proceeds as follows. At each time t, upon arrival of ball t, two bins are picked uniformly at random. If both bins are honest, then the ball lands in the bin with fewer balls (ties are solved arbitrarily). If at least one bin is faulty, then a malicious adversary decides which bin receives the ball and is capable to enforce it on the algorithm. See also Algorithm 1.

This is formally done as follows: the adversary controlling the faulty bin could provide the algorithm with an arbitrary number (instead of the actual number of balls in that bin), which is then compared with the number provided by the other bin (potentially honest) and the minimum is selected. We assume that the adversary has the following attributes:

- it is *Byzantine*, i.e., it can provide an arbitrary answer about the number of balls in the selected faulty bin(s),
- it is *computationally unbounded*,
- it *knows the algorithm* BT_PoTC,

– it is *strongly adaptive*, i.e., it knows which bins have been randomly selected so far and which bins will be selected in the future and can use this knowledge to provide a potentially malicious answer to the algorithm about the number of balls in the currently selected faulty bin(s).

We measure the maximum number of balls in honest bins, and the goal is to have as precise asymptotic upper bound on this number as possible. This estimate is supposed to hold *with high probability*, which denotes probability $1-o(1)$. Part of the analysis is done with high probability that is polynomially close to 1, i.e., $1 - n^{-c}$ for some constant $c > 0$. Throughout the paper we use asymptotic notation of $O(\cdot)$ and $o(\cdot)$ to upper bound formulas and (complementary) probabilities: $f(x) = O(g(x))$ means that $f(x) \le c \cdot g(x)$ for some sufficiently large constant c, while $f(x) = o(g(x))$ means that $f(x)/g(x)$ converges to 0 when x goes to infinity.

Challenges. At first glance it may seem that one can directly apply the results about PoTC process to this problem. E.g., we split the balls into those that choose between 2 honest bins and those that choose between an honest and a faulty bin. The balls that choose between 2 honest bins behave like PoTC process, while the balls that choose between an honest and a faulty bin behave like the standard balls-into-bins process. However, those two processes are dependent on each other – the balls added by the latter (standard balls-into-bins) process affect decisions made during the power-of-two-choices process. This difficulty is not easily worked around. For example, a related problem with similar dependencies required a completely different, non-trivial method of analysis [26].

2.2 Our results

First we present a result for $m > \frac{3}{2} n \log n$. Note that this result covers any number of faulty bins $f < n$. The proof is given in Section 5.

Theorem 1. *For any values of parameters $m, \delta \in (0,1)$ and $c > 1$ such that $\frac{2m\delta^2}{3n} > c \log n$, there is no honest bin with more than $(1+\delta)\frac{2m}{n}$ balls in it with probability at least $1 - n^{1-c}$.*

E.g., with $m > 9n \log n$, $\delta = 1/2$ and $c = \frac{3}{2}$, we get the following corollary.

Corollary 1. *Let $m > 9n \log n$. Then, there is no honest bin with more than $\frac{3m}{n}$ balls in it with probability at least $1 - n^{-1/2}$.*

Now we present a result complimentary with Theorem 1, i.e., a result that works for $m \le \frac{3}{2} n \log n$. This is the main result of this article, and its proof is given in Section 3.

Theorem 2. *Let $m = O(n \log n)$ and $1 \le f < \frac{n^2}{2em}$. Let ϕ be such that the number of honest bins with at least ϕ balls is at most $\frac{n^2}{4em}$ with probability at least $1 - n^{-2}$. Then, there is no honest bin with more than*

$$\phi + \log\log \frac{n^2}{2emf} + \log_{\frac{n^2}{2emf}} \frac{2f}{\sqrt{f^2 + 8\frac{n^2}{m}\log n} - f} + 1$$

balls in it, with probability at least $1 - o(1)$.

Next we present corollaries with clearer formulas. We have:

$$
\begin{aligned}
O\left(\log_{\frac{n^2}{2emf}} \frac{f}{\sqrt{f^2 + \frac{n^2}{m}\log n} - f}\right) &\le O\left(\log_{\frac{n^2}{2emf}} \frac{f \cdot \left(\sqrt{f^2 + \frac{n^2}{m}\log n} + f\right)}{f^2 + \frac{n^2}{m}\log n - f^2}\right) \\
\le O\left(\log_{\frac{n^2}{2emf}} \frac{f^2 + f\sqrt{\frac{n^2}{m}\log n}}{\frac{n^2}{m}\log n}\right) &\le O\left(\log_{\frac{n^2}{2emf}} \left(\frac{f^2 m}{n^2 \log n} + \sqrt{\frac{f^2 m}{n^2 \log n}}\right)\right) \\
\le O\left(\max\left\{\log_{\frac{n^2}{2emf}} \left(\frac{2emf}{n^2} \cdot \frac{f}{2e\log n}\right), \log_{\frac{n^2}{2emf}} \left(\sqrt{\frac{2emf}{n^2} \cdot \frac{f}{2e\log n}}\right)\right\}\right) \\
\le O\left(\log_{\frac{n^2}{2emf}} \frac{f}{2e\log n}\right) . \qquad (1)
\end{aligned}
$$

We start with a corollary that is not yet fully optimized, and finish with the one in optimal form.

Corollary 2. *Let* $1 \le f < \frac{n^2}{2em}$. *Then, there is no honest bin with more than*

$$
O\left(\left(\frac{m}{n}\right)^2 + \log\log \frac{n^2}{2emf} + \log_{\frac{n^2}{2emf}} \frac{f}{2e\log n}\right)
$$

balls in it, with probability at least $1 - o(1)$.

Proof. Let $\phi = \frac{4em^2}{n^2}$. Note that the number of honest bins with at least ϕ balls is at most $m/\phi = \frac{n^2}{4em}$ with probability 1. Therefore, we can use Theorem 2 to obtain the corollary, bounding the last logarithm as in Equation (1).

Corollary 3. *Let* $1 \le f < \frac{n^2}{2em}$ *and* $m = O(n)$. *There is no honest bin with more than*

$$
O\left(\log\log \frac{n^2}{mf} + \log_{\frac{n^2}{2emf}} \frac{f}{2e\log n}\right)
$$

balls in it, with probability at least $1 - o(1)$.

Proof. The result follows directly from Corollary 2 and bound $m = O(n)$.

The following corollary uses the fact that the number of honest bins with at least $\phi = \Theta(m/n)$ balls in them is at most $\beta_\phi = \frac{n^2}{4em}$, for $m > en$ and $f < \frac{n^2}{2em}$. This is formally shown in Lemma 7 in Section 4.

Corollary 4. *Let* $1 \le f < \frac{n^2}{2em}$, *where* $m \ge en$ *and* $m = O(n\log n)$. *There is no honest bin with more than*

$$
O\left(\frac{m}{n} + \log\log \frac{n^2}{mf} + \log_{\frac{n^2}{2emf}} \frac{f}{2e\log n}\right)
$$

balls in it, with probability at least $1 - o(1)$.

Proof. Let $\phi = \frac{cem}{n}$ for some $c > 2e$. According to Lemma 7, the number of honest bins with at least ϕ balls is at most $\beta_\phi = \frac{n^2}{4em}$ (which is denoted by event $\mathcal{E}_\phi$) with probability $1 - n^{-c}$. Therefore, we can use Theorem 2 to obtain the corollary, bounding the last logarithm as in Equation (1). (The detailed definitions of β_ϕ and $\mathcal{E}_\phi$ are given below in Subsection 3.1.)

Corollary 5. *Let* $1 \leq f < n \cdot \max\{\frac{1}{12e \log n}, \frac{n}{2em}\}$*, where* $m \geq n$*. There is no honest bin with more than*

$$O\left(\frac{m}{n} + \log\log\frac{n^2}{mf} + \log_{\frac{n^2}{2emf}}\frac{f}{2e\log n}\right) \leq O\left(\frac{m}{n} + \log\log\frac{n}{f} + \log f\right)$$

balls in it, with probability at least $1 - o(1)$*.*

Proof. For $m \geq 6n \log n$, the formula is just $O(m/n)$, by Theorem 1. Consider $en \leq m < 6n \log n$. Observe that $\frac{n}{6f \log n} < \frac{n^2}{mf} \leq \frac{n}{ef}$, therefore we upper bound the formula in Corollary 4 as follows:

$$\begin{aligned} O\left(\log\log\frac{n^2}{mf} + \log_{\frac{n^2}{2emf}}\frac{f}{2e\log n}\right) &\leq O\left(\log\log\frac{n}{f} + \frac{\log\frac{f}{2e\log n}}{\log\frac{n}{f\log n}}\right) \\ &\leq O\left(\log\log\frac{n}{f} + \log f\right) . \end{aligned}$$

Note that all the results for $(1+\beta)$-choice process by Peres et al. [26] are asymptotically at least $\log n$ w.h.p. On the other hand, our analysis provides a tighter bound on the maximum load of a bin for some parameters, e.g., for $m = O(n)$ and $f = o(n)$ we can get from Corollary 5 that the maximum load is $O(\log\log(n/f) + \log f)$ w.h.p.

3 Probabilistic Analysis of Power of Two Choices under Byzantine Faults — Proof of Theorem 2

3.1 Overview of the analysis

Our analysis is inspired by [5]. While the general idea of the proof still applies, introduction of faulty bins controlled by a strong Byzantine adversary called for additional more careful analysis to be made. Another challenge lies in finding appropriate values of parameters that take into account malicious influence of faulty bins that may interact with honest bins (when selected together by the random process).

The main effort is in analyzing events $\mathcal{E}_i$, representing that there exist at most β_i honest bins with at least i balls in each, for a range of positive integers i. Values β_i will be defined later.

The starting point of our analysis is for some carefully chosen threshold parameter $i = \phi$ such that the event $\mathcal{E}_\phi$ is guaranteed to happen with high probability. (We prove later in Section 4 and Lemma 7 that such $\phi = O(m/n)$, achieving asymptotically perfect allocation threshold, exists and thus could be substituted in the formula of Theorem 2 to obtain an efficient bound in Corollary 4.) As we move to larger parameters $i > \phi$, the events $\mathcal{E}_i$ represent situations where there are fewer and fewer honest bins with more and more balls in them. Eventually, we reach a final event $\mathcal{E}_k$ such that there is less than 1 bin with at least k balls in it, for some value of k to be derived in the analysis; meaning – there are no such bins.

We analyze the probability that each of the events $\mathcal{E}_i$ occurs. The starting point $\mathcal{E}_\phi$ is guaranteed with high probability. We want to show that all the events $\mathcal{E}_i$ occur with high probability until *after* the event we need, $\mathcal{E}_k$. However, the approach we use can only show that $\mathcal{E}_{k-2}$ occurs with high probability. Still, having event $\mathcal{E}_{k-2}$, we can manually show that the number of bins with at least k balls is less than 1.

The proof of Theorem 2 is divided into four steps. Recall that in this analysis we assume $m = O(n \log n)$, while the complementary case is analyzed in Section 5 and Theorem 1.

In Step 1, we define critical sequences of parameters β_i and events $\mathcal{E}_i$, taking into account the impact of faulty bins. We obtain the initial estimate of the probability $P(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i)$. Here $\neg\mathcal{E}_{i+1}$ denotes the event complementary to $\mathcal{E}_{i+1}$, i.e., the event that $\mathcal{E}_{i+1}$ does not hold.

In Step 2, we find that, with an additional assumption for the threshold parameter i, event $\mathcal{E}_i$ occurs with high probability.

In Step 3, we calculate for which values of i the assumption considered in Step 2 actually holds. As long as the assumption holds, the good events $\mathcal{E}_i$ hold too. In this step, the impact of faulty bins on the analysis, and their number f in the formulas, is the most challenging to deal with.

In Step 4, we use the highest event $\mathcal{E}_i$ in the sequence that we could prove to hold with high probability. Instead of continuing to show that further good events in the sequence hold with high probability, we prove directly that the number of bins with many balls is less than 1, which concludes the proof of Theorem 2.

3.2 Worst-case adversary

First, we make an observation about the worst case adversary.

Consider a greedy adversary $GREEDY$ that, whenever a ball chooses between a faulty bin and an honest bin, makes the ball land in the honest bin. We claim that such a greedy adversary is the worst-case adversary even among the strongly adaptive adversaries.

Lemma 1. *$GREEDY$ is the worst-case adversary for maximizing load.*

Proof idea. Consider any adversary ADV. Consider any bin b and any sequence of exchanges S. Let $l_b^A(S,t)$ be the load of bin b at time t under adversary ADV for the sequence of exchanges S. A simple inductive argument over time t shows that – for every honest bin b, all sequences of exchanges S and at any time t – we have $l_b^{ADV}(S,t) \leq l_b^{GREEDY}(S,t)$.

n	The number of bins.
m	The total number of balls that will be thrown into the bins.
f	The number of faulty bins.
e	Euler's number, $e = 2.71828\ldots$
i	Parameter of the considered sequence of events $\{\mathcal{E}_i\}$, in which i corresponds to a threshold on the load of the bins.
t	Actually considered time step and ball number.
v_i^t	The number of bins that are honest and have at least i balls at time t.
h^t	The number of balls in the bin in which the t-th ball has landed at time t (measured after it has landed).
u_i^t	The number of balls in honest bins that have height at least i at time t.
β_i	Upper bound on the number of honest bins with at least i balls.
$\mathcal{E}_i$	An event such that $v_i^m \leq \beta_i$.
ϕ	The starting point for the considered sequence of events $\mathcal{E}_i$. $\beta_\phi = \frac{n^2}{4em}$ and event $\mathcal{E}_\phi$ occurs with high probability.
Y_i^t	$Y_i^t = 1$ if the t-th ball landed in an honest bin and $h^t \geq i+1$ and $v_i^{t-1} \leq \beta_i$; otherwise, $Y_i^t = 0$.
p_i	Upper bound on probability that a ball landed in a bin with at least i balls in it. We have $p_i = \left(\frac{\beta_i}{n}\right)^2 + \frac{f}{n} \cdot \frac{\beta_i}{n}$.
i_1	$\phi + i_1$ is the smallest index of β_i such that $\beta_{\phi+i_1} < f$.
i^*	The smallest value such that $mp_{\phi+i^*} < 2\ln n$.
$B(m,p)$	Binomial distribution with m trials and probability p.

Table 2. Table of problem parameters and most important notation in the analysis.

The full proof is deferred to the full version of the paper.

From now on, we analyze the protocol against the $GREEDY$ adversary. According to Lemma 1, the results will also apply to an arbitrary adversary.

3.3 Step 1: initial estimate of $P(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i)$

The goal of the first step of the analysis is to obtain the initial estimate on $P(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i)$ for any integer $i > 0$.

In our analysis we will use the following two standard facts. Let $B(m,p)$ denote the binomial distribution with m trials and probability of success p. The following stochastic dominance has been proved and used before in various probabilistic analyses, c.f. [14, in the analysis of Lemma 2].

Fact 1 *Let $X_1, X_2, \ldots, X_m$ be a sequence of random variables with values in an arbitrary domain, and let $Y_1, Y_2, \ldots, Y_m$ be a sequence of binary random variables, with the property that $Y_i = Y_i(X_1, \ldots, X_i)$.*

If $P(Y_i = 1 \mid X_1, \ldots, X_{i-1}) \leq p$ then $P\left(\sum_{i=1}^m Y_i \geq k\right) \leq P(B(m,p) \geq k)$.

The proof of the next fact (Chernoff Bounds) can be found, e.g., in [19].

Fact 2 (Chernoff Bounds) *For $a \geq mp$, $P(B(m,p) \geq a) \leq \left(\frac{mp}{a}\right)^a e^{a-mp}$. If $a = (1+\delta)mp$, for some $\delta \in (0,1)$, then $P(B(m,p) \geq a) \leq e^{-mp\delta^2/3}$.*

We start the technical analysis by defining three crucial notations: $v_i^t, \mathcal{E}_i, \beta_i$. Here, an integer parameter i denotes the considered threshold on the number of balls in bins, and t stands for the order number of analyzed ball. The list of all important notations and their meaning is given in Table 2. Let v_i^t be the number of bins that are honest and have at least i balls at time t, i.e., after considering and placing t balls. Let $\mathcal{E}_i$ be an event such that $v_i^m \leq \beta_i$, for given parameters β_i that will be determined later. Let index ϕ be such that $\mathcal{E}_\phi$ occurs with probability at least $1 - n^{-2}$.

Our goal is to find parameters β_i such that $\mathcal{E}_i$ holds with high probability and $\beta_i < 1$ for as small parameter i as possible.

Let the *height of ball* t, denoted h^t, be the number of balls in the bin in which the t-th ball has landed (measured after it has landed).

Let Y_i^t be a random variable such that $Y_i^t = 1$ if the t-th ball landed in an honest bin and $h^t \geq i+1$ and $v_i^{t-1} \leq \beta_i$; otherwise, $Y_i^t = 0$. Intuitively, $Y_i^t = 1$ denotes the event such that a ball landed in a bin with already many balls in it. The additional constraint $v_i^{t-1} \leq \beta_i$ helps to carry on the analysis leading to finding a small value i for which $\beta_i < 1$ and $\mathcal{E}_i$ holds with high probability.

Let ω_j be a random variable equal to the bin number where the j-th ball has landed. We will upper bound the probability $P(Y_i^t = 1 \mid \omega_1, \omega_2, \ldots, \omega_{t-1})$. Note that $Y_i^t = 1$ only if one of the two cases occurs:

- the protocol picks two honest bins with at least i balls in them to choose from, which takes place with probability at most $\left(\frac{\beta_i}{n}\right)^2$, or
- if one chosen bin is faulty and the other is honest, with at least i balls, which happens with probability at most $\frac{f}{n} \cdot \frac{\beta_i}{n}$.

Therefore, we get the following inequality

$$P(Y_i^t = 1 \mid \omega_1, \omega_2, \ldots, \omega_{t-1}) \leq \left(\frac{\beta_i}{n}\right)^2 + \frac{f}{n} \cdot \frac{\beta_i}{n} \ . \tag{2}$$

Let $p_i = \left(\frac{\beta_i}{n}\right)^2 + \frac{f}{n} \cdot \frac{\beta_i}{n}$.

For parameters i and t, let u_i^t be the number of balls in honest bins that have height at least i at time t. Now we will relate u_{i+1}^m with random variables Y_i^t. Note that, if $\mathcal{E}_i$ holds, then $v_i^{t-1} \leq \beta_i$ for all t. In that case $\sum_{t=1}^m Y_i^t = u_{i+1}^m$. Therefore, $P(\sum_{t=1}^m Y_i^t \geq k \mid \mathcal{E}_i) = P(u_{i+1}^m \geq k \mid \mathcal{E}_i)$ for any parameter k.

Now we will estimate $P(v_{i+1}^m \geq k \mid \mathcal{E}_i)$ for some values of parameter k. Note that $v_i^t \leq u_i^t$ for all i and t, and therefore

$$P(v_{i+1}^m \geq k \,|\, \mathcal{E}_i) \leq P(u_{i+1}^m \geq k \,|\, \mathcal{E}_i) = P\left(\sum_{t=1}^m Y_i^t \geq k \mid \mathcal{E}_i\right) \leq \frac{P(B(m,p_i) \geq k)}{P(\mathcal{E}_i)} \ ,$$

where the last inequality follows from Fact 1.

For $k = \beta_{i+1}$ we get

$$P(v_{i+1}^m \geq \beta_{i+1} \mid \mathcal{E}_i) \leq \frac{P(B(m,p_i) \geq \beta_{i+1})}{P(\mathcal{E}_i)} \ . \tag{3}$$

We can use Fact 2 with $a = \beta_{i+1}$, where $\beta_{i+1} = emp_i$ and get

$$P(v_{i+1}^m \geq \beta_{i+1} \mid \mathcal{E}_i) \leq \left(\frac{1}{e}\right)^{emp_i} e^{(e-1)mp_i} = e^{-mp_i} \ .$$

Note that the event $(v_{i+1}^m > \beta_{i+1})$ is equivalent to $\neg\mathcal{E}_{i+1}$. Therefore,

$$P(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \leq e^{-mp_i} \ . \tag{4}$$

3.4 Step 2: analysis of event $\mathcal{E}_i$ for i corresponding to $p_i \geq \frac{2\ln n}{m}$

In Step 2 we use the estimate obtained in the previous step to establish that events $\mathcal{E}_i$ hold with high probability (polynomially close to 1) for some parameters i. We will analyze later in Step 3 which values of i hold this property.

Our goal now is to show that $P(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \leq 1/n^2$. Assume we have some ϕ such that $\mathcal{E}_\phi$ occurs with high probability for $\beta_\phi = \frac{n^2}{4em}$ – existence of such ϕ will be shown later in Lemma 7. If this is the case, then $\mathcal{E}_i$ holds for all $\phi \leq i \leq \phi + i^*$ for some i^*, also with high probability.

Suppose first that $mp_i \geq 2\ln n$. Then $e^{-mp_i} \leq 1/n^2$ and, by Equation 4, we have $P(\neg\mathcal{E}_{i+1}|\mathcal{E}_i) \leq 1/n^2$. Then, we could follow a simple inductive argument to prove Lemma 2 below. The proof is deferred to the full version of the paper.

Lemma 2. *If $mp_i \geq 2\ln n$ holds for consecutive parameters i starting from ϕ, then $P(\neg\mathcal{E}_\phi \vee \neg\mathcal{E}_{\phi+1} \vee \cdots \vee \neg\mathcal{E}_{\phi+i}) \leq (i+1)/n^2$.*

3.5 Step 3: finding i^* – the minimum i satisfying $p_{\phi+i} < \frac{2\ln n}{m}$

The previous step proved that events $\mathcal{E}_i$ hold with high probability for i such that $mp_i \geq 2\ln n$. In this step we answer the question: What are the values of i such that the sought property $mp_i \geq 2\ln n$ holds? To address it, we will now find the minimum i^* such that $mp_{\phi+i^*} < 2\ln n$.

Lemma 3. *The smallest value i^* such that $mp_{\phi+i^*} < 2\ln n$ satisfies*

$$i^* \leq \log_2 \log_{\frac{2em\beta_\phi}{n^2}} \frac{2emf}{n^2} \ + \ \log_{\frac{n^2}{2emf}} \frac{2f}{-f + \sqrt{f^2 + 8\frac{n^2}{m}\log n}} \ . \tag{5}$$

In order to prove Lemma 3, we introduce Lemmas 4–6. The first of them, Lemma 4, can be proved by induction on parameter i. The proof of Lemma 4 is deferred to the full version of the paper.

Lemma 4. *$\beta_{\phi+i}$ is monotonically decreasing when parameter i increases.*

Lemma 5. *Let i_1 be the smallest i such that $\beta_{\phi+i} < f$. It holds that $i_1 \leq \log_2 \log_{\frac{2em\beta_\phi}{n^2}} \frac{2emf}{n^2}$.*

Proof. Consider values of parameter j such that $\beta_j \geq f$.

We get that $p_i \leq 2\left(\frac{\beta_j}{n}\right)^2$ and the following recursive equation $\beta_{j+1} \leq em \cdot 2\left(\frac{\beta_j}{n}\right)^2$. Therefore, $\beta_{j+i} \leq \frac{n^2}{2em}\left(\frac{2em\beta_j}{n^2}\right)^{2^i}$ for i such that $\beta_j \geq f$, $\beta_{j+1} \geq f$, …, $\beta_{j+i-1} \geq f$. In particular, we get

$$\beta_{\phi+i} \leq \frac{n^2}{2em}\left(\frac{2em\beta_\phi}{n^2}\right)^{2^i} \tag{6}$$

for i such that $\beta_{\phi+i-1} \geq f$ (it follows that $\beta_\phi \geq f$, $\beta_{\phi+1} \geq f$, …, $\beta_{\phi+i-2} \geq f$ according to Lemma 4).

Now we look at what values of i are such that $\beta_{\phi+i-1} \geq f$. We can look for the smallest i_1 such that the opposite occurs, i.e., $\beta_{\phi+i_1} < f$ (recall that we only consider $f \geq 1$ – see the statement of Theorem 2). Recall our choice $\beta_\phi = \frac{n^2}{4em} < \frac{n^2}{2em}$. Note the base of the power in Inequality 6 is $\frac{2em\beta_\phi}{n^2} < 1$. Therefore, we can find that $\beta_{\phi+i} < f$ for any $i \geq \log_2 \log_{\frac{2em\beta_\phi}{n^2}} \frac{2emf}{n^2}$ (note that $\log_{\frac{2em\beta_\phi}{n^2}} \frac{2emf}{n^2} > 0$, since $\frac{2em\beta_\phi}{n^2} = \frac{1}{2}$ and $\frac{2emf}{n^2} < 1$ due to our assumption that $f < \frac{n^2}{2em}$ in the statement of Theorem 2). It follows that the smallest i_1 such that $\beta_{\phi+i_1} < f$ satisfies: $i_1 \leq \log_2 \log_{\frac{2em\beta_\phi}{n^2}} \frac{2emf}{n^2}$.

Note that $\beta_{\phi+i_1}$ is the first β such that $\beta_{\phi+i_1} < f$.

Lemma 6. *For $i \geq 0$, $\beta_{\phi+i_1+i} \leq \left(\frac{2emf}{n^2}\right)^i \beta_{\phi+i_1}$.*

Proof. According to Lemma 5 and Lemma 4, we have $\beta_j < f$ for all $j \geq \phi + i_1$. We get $\beta_{j+1} < em\left(\frac{f}{n} \cdot \frac{\beta_j}{n} + \frac{f}{n} \cdot \frac{\beta_j}{n}\right) = \frac{2emf}{n^2}\beta_j$. Therefore, $\beta_{j+i} \leq \left(\frac{2emf}{n^2}\right)^i \beta_j$ for i, j such that $\beta_j < f$, $\beta_{j+1} < f$, …, $\beta_{j+i-1} < f$. In particular, for any $i \geq 0$,

$$\beta_{\phi+i_1+i} \leq \left(\frac{2emf}{n^2}\right)^i \beta_{\phi+i_1} \ . \tag{7}$$

Now we are ready to prove Lemma 3.

Proof (Proof of Lemma 3).

First, recall that $\beta_{i+1} = emp_i$ and $p_i = \left(\frac{\beta_i}{n}\right)^2 + \frac{f}{n} \cdot \frac{\beta_i}{n}$ for all i.

We are interested in i such that $mp_{\phi+i} < 2 \ln n$, that is to say

$$\left(\frac{\beta_{\phi+i}}{n}\right)^2 + \frac{f}{n} \cdot \frac{\beta_{\phi+i}}{n} < \frac{2 \ln n}{m} \ . \tag{8}$$

We can treat this inequality as a quadratic inequality with variable $\beta_{\phi+i}$ to obtain that $mp_{\phi+i} < 2 \ln n$ holds if

$$\beta_{\phi+i} \in \left(\frac{-f - \sqrt{f^2 + 8\frac{n^2}{m}\log n}}{2}, \frac{-f + \sqrt{f^2 + 8\frac{n^2}{m}\log n}}{2}\right) \ . \tag{9}$$

Because of the monotonicity of $\beta_{\phi+i}$, as in Lemma 4, we are interested in the smallest i such that

$$\beta_{\phi+i} < \frac{-f + \sqrt{f^2 + 8\frac{n^2}{m}\log n}}{2} \ . \tag{10}$$

Based on Lemma 6 and the facts that $\beta_{\phi+i_1} < f$ (see Lemma 5) and $f < \frac{n^2}{2em}$ (assumption in Theorem 2), we get that $\beta_{\phi+i_1+i} < \frac{-f+\sqrt{f^2+8\frac{n^2}{m}\log n}}{2}$ holds for any $i \geq \log_{\frac{2emf}{n^2}} \frac{-f+\sqrt{f^2+8\frac{n^2}{m}\log n}}{2f}$. It follows that $i^* \leq i_1 + \log_{\frac{n^2}{2emf}} \frac{2f}{-f+\sqrt{f^2+8\frac{n^2}{m}\log n}}$.

3.6 Step 4: finalizing the analysis - beyond threshold parameter $\phi+i^*$

Now that we have the value of i^* and an analysis of events $\mathcal{E}_i$, for $i \leq i^*$, we can find the probability that there exists an honest bin with more than $\phi + i^*$ balls in it. The result is that there are no honest bins with at least $\phi + i^* + 2$ balls, with high probability of $1 - o(1)$.

Recall that $\beta_{i+1} = emp_i$ for all i. In particular, since $mp_{\phi+i^*} < 2\ln n$ we have:

$$\beta_{\phi+i^*+1} = emp_{\phi+i^*} < 2e\ln n \ . \tag{11}$$

Consider $v^m_{\phi+i^*+1}$. We get

$$P(v^m_{\phi+i^*+1} \geq 2e\ln n \mid \mathcal{E}_{\phi+i^*}) \quad \leq \quad \frac{P\left(B(m, p_{\phi+i^*}) \geq 2e\ln n\right)}{P(\mathcal{E}_{\phi+i^*})} \tag{12}$$

$$< \quad \frac{P\left(B\left(m, \frac{2\ln n}{m}\right) \geq 2e\ln n\right)}{P(\mathcal{E}_{\phi+i^*})} \quad \leq \quad \frac{1}{n^2 P(\mathcal{E}_{\phi+i^*})} \ , \tag{13}$$

where the last inequality follows from Fact 2.

Finally, we can bound $P(u^m_{\phi+i^*+2} < 1)$ in the following derivation.

$$P(u^m_{\phi+i^*+2} \geq 1 \mid v^m_{\phi+i^*+1} < 2e\ln n) \leq \frac{P\left(B\left(m, \left(\frac{2e\ln n}{n}\right)^2 + \frac{f}{n}\frac{2e\ln n}{n}\right) \geq 1\right)}{P(v^m_{\phi+i^*+1} < 2e\ln n)} \tag{14}$$

$$\leq \frac{m\left(\left(\frac{2e\ln n}{n}\right)^2 + \frac{f}{n}\frac{2e\ln n}{n}\right)}{P(v^m_{\phi+i^*+1} < 2e\ln n)} \ . \tag{15}$$

Fact 3 *For any events A, B, we have: $P(A) \leq P(A|B) \cdot P(B) + P(\neg B)$.*

Using Fact 3 twice, we get

$$\begin{aligned} P(u^m_{\phi+i^*+2} \geq 1) &\leq P(u^m_{\phi+i^*+2} \geq 1 \mid u^m_{\phi+i^*+1} < 2e\ln n) \cdot P(u^m_{\phi+i^*+1} < 2e\ln n) + \\ &\quad + P(u^m_{\phi+i^*+1} \geq 2e\ln n) \\ &\leq P(u^m_{\phi+i^*+2} \geq 1 \mid u^m_{\phi+i^*+1} < 2e\ln n) \cdot P(u^m_{\phi+i^*+1} < 2e\ln n) + \\ &\quad + P(u^m_{\phi+i^*+1} \geq 2e\ln n \mid \mathcal{E}_{\phi+i^*}) \cdot P(\mathcal{E}_{\phi+i^*}) + P(\neg\mathcal{E}_{\phi+i^*}) \ . \end{aligned}$$

Using Equations 13 and 15, we get

$$P(u^m_{\phi+i^*+2} \geq 1) \leq m\left(\left(\frac{2e\ln n}{n}\right)^2 + \frac{f}{n}\frac{2e\ln n}{n}\right) + \frac{1}{n^2} + P(\neg\mathcal{E}_{\phi+i^*})$$
$$\leq m\left(\left(\frac{2e\ln n}{n}\right)^2 + \frac{f}{n}\frac{2e\ln n}{n}\right) + \frac{1}{n^2} + P(\neg\mathcal{E}_{\phi+1} \vee \cdots \vee \neg\mathcal{E}_{\phi+i^*})$$
$$\leq O\left(\frac{\ln^3 n}{n} + \frac{f\ln^2 n}{n}\right) + \frac{1}{n^2} + \frac{i^*+1}{n^2} ,$$

where the last line follows from Lemma 2 and the assumed bound $m = O(n\log n)$.

For $f = o\left(\frac{n}{\ln^2 n}\right)$, we get

$$P(u^m_{\phi+i^*+2} \geq 1) = o(1) . \tag{16}$$

Finally, this means that $u^m_{\phi+i^*+2} < 1$ occurs with probability $1 - o(1)$. Therefore, there are no balls with height at least $\phi + i^* + 2$ with high probability, which also means that there are no honest bins with at least $\phi + i^* + 2$ balls in them which completes the proof of Theorem 2.

4 Case $m \geq en$ – Setting up Initial Value of ϕ

Lemma 7. *Assume $m \geq en$ and $f < \frac{n^2}{2em}$. Let $\phi = c\frac{em}{n}$ and $\beta_\phi = \frac{n^2}{4em}$, for some constant $c > 2e$ to be defined later in the proof. Then $\mathcal{E}_\phi$ holds with probability at least $1 - n^{-c}$.*

Proof. We first compute an upper bound on the probability of the complementary event, that is, that there is a subset of $\beta_\phi + 1$ of honest bins containing at least ϕ balls each. It is upper bounded by a union of events, parameterized by any subset B of honest bins, that each bin in B has at least ϕ balls. This can be further upper bounded by multiplying the number of such sets B, $\binom{n-f}{\beta_\phi+1}$, by the product of upper bounds on the probability that an y-th bin in B, where $0 \leq y \leq \beta_\phi$ was randomly selected at least ϕ times when allocating the remaining at least $m - y\phi$ balls, $\binom{m-y\phi}{\phi}\left(\frac{2-\frac{1}{n}}{n}\right)^\phi$. This results in the following upper bound formula and its further transformation:

$$\binom{n-f}{\beta_\phi+1}\prod_{y=0}^{\beta_\phi}\left(\binom{m-y\phi}{\phi}\left(\frac{2-\frac{1}{n}}{n}\right)^\phi\right)$$

$$= \binom{n-f}{\frac{n^2}{4em}+1}\prod_{y=0}^{\frac{n^2}{4em}}\left(\binom{m-yc\frac{em}{n}}{c\frac{em}{n}}\left(\frac{2-\frac{1}{n}}{n}\right)^{c\frac{em}{n}}\right)$$

$$\leq 2^{n-f}\left(\left(\frac{me}{c\frac{em}{n}}\right)^{c\frac{em}{n}}\left(\frac{2}{n}\right)^{c\frac{em}{n}}\right)^{\frac{n^2}{4em}+1} = 2^{n-f}\left(\frac{2}{c}\right)^{c\frac{n}{4}+c\frac{em}{n}} \leq 2^{n-f}\left(\frac{2}{c}\right)^{c\frac{n}{4}} ,$$

where the first inequality follows from bounds $2 - \frac{1}{n} \leq 2$, $\binom{m - yc\frac{em}{n}}{c\frac{em}{n}} \leq \binom{m}{c\frac{em}{n}}$ for $0 \leq y \leq \beta_\phi$, and $\binom{n}{x} \leq (ne/x)^x$; the second inequality follows from the assumption $c > 2e$ and from the monotonicity of the exponent function. Next, observe that

$$2^{n-f}\left(\frac{2}{c}\right)^{\frac{cn}{4}} = \exp\left((n-f)\ln 2 - \frac{cn}{4}\ln\frac{c}{2}\right) < n^{-c}$$

for any c satisfying $\frac{cn}{4}\ln\frac{c}{2} - c\ln n > (n-f)\ln 2$, for instance, for $c \geq 2e^4$ and any $n \geq 3$,[4] or for any $c > 2e$ and any sufficiently large n.

5 Case $m > \frac{3}{2}n\log n$ – Proof of Theorem 1

In this section we prove Theorem 1. Consider an honest bin j. For any $t \leq m$, let X_t be a random variable equal to 1 if ball t lands in bin j, and equal to 0 otherwise; let Y_t be equal to 1 if bin j is randomly selected at time t, and equal to 0 otherwise. Observe that:

$$\forall_{t\leq m} X_t \leq Y_t \qquad \text{and} \qquad \forall_{t\leq m} P(X_t = 1) \leq P(Y_t = 1) = \frac{2}{n}\ .$$

Let $\mu_X = \mathbf{E}\left[\sum_{t\leq m} X_t\right]$ and $\mu_Y = \mathbf{E}\left[\sum_{t\leq m} Y_t\right] = \frac{2m}{n}$. By Fact 2, for any $\delta \in (0,1)$:

$$P\left(\sum_{t\leq m} X_t > (1+\delta)\mu_Y\right) \leq P\left(\sum_{t\leq m} Y_t > (1+\delta)\mu_Y\right) < e^{\frac{-\mu_Y\delta^2}{3}} = e^{-\frac{2m\delta^2}{3n}}\ .$$

Assuming that $\frac{2m\delta^2}{3n} > c\log n$, for some $c > 1$, which holds for some $\delta \in (0,1)$ and some $c > 1$ as long as $m > \frac{3}{2}n\log n$, we get $P\left(\sum_{t\leq m} X_t > (1+\delta)\frac{2m}{n}\right) < n^{-c}$.

Consequently, by applying the union bound over all bins, the probability that there is a bin with more than $(1+\delta)\frac{2m}{n}$ in it is at most n^{1-c}.

6 Discussion and Open Problems

In this work, we provided an analysis of the efficiency of the popular load balancing rule – the Power of Two Choices – in the system that some bins are controlled by a malicious adversary. For $m = O(n\log n)$, we showed that the maximum load on any honest bin has a logarithmic dependence on the number of faulty nodes f.

There are several open questions to be explored. One open question is related to the tightness of our bounds. It will be interesting to obtain matching lower

[4] Observe that the power of two choices makes use of randomness only for $n \geq 3$ bins.

bounds related to the number of faults f, and also m and n. It will also be interesting to explore the graphical case. In the case without faults, the problem has been studied for regular graphs and the maximum load depends on the node degree. It will be interesting to explore the dependence of the load on the number of faulty graph nodes f and the degree of the regular graph.

Another open problem is the lower bound on the minimum load of a bin as well as the gap between the maximum and minimum loads. These bounds would be used to prove the Fault-Tolerant Phase Clock works in population protocols.

Acknowledgments. This study was partly supported by the National Science Center, Poland (NCN), grant 2020/39/B/ST6/03288, and by the (US) National Science Foundation grant CNS-2131538.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdulazeez, M., Garncarek, P., Kowalski, D.R., Wong, P.W.H.: Lightweight robust framework for workload scheduling in clouds. In: IEEE International Conference on Edge Computing, EDGE 2017. pp. 206–209. IEEE Computer Society (2017)
2. Alistarh, D., Aspnes, J., Gelashvili, R.: Space-optimal majority in population protocols. In: Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). pp. 2221–2239 (2018)
3. Angluin, D., Aspnes, J., Diamadi, Z., Fischer, M.J., Peralta, R.: Computation in networks of passively mobile finite-state sensors. Distributed Comput. **18**(4), 235–253 (2006)
4. Angluin, D., Aspnes, J., Eisenstat, D.: A simple population protocol for fast robust approximate majority. Distributed Comput. **21**(2), 87–102 (2008)
5. Azar, Y., Broder, A.Z., Karlin, A.R., Upfal, E.: Balanced allocations. SIAM Journal on Computing **29**(1), 180–200 (1999)
6. Bansal, N., Feldheim, O.N.: The power of two choices in graphical allocation. In: STOC 2022. p. 52–63. ACM (2022)
7. Bansal, N., Kuszmaul, W.: Balanced allocations: The heavily loaded case with deletions. In: 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS). pp. 801–812 (2022)
8. Ben-Nun, S., Kopelowitz, T., Kraus, M., Porat, E.: An $O(\log^{3/2} n)$ parallel time population protocol for majority with $O(\log n)$ states. In: Proc. 39th Symposium on Principles of Distributed Computing. p. 191–199. PODC '20, ACM (2020)
9. Ben-Or, M., Pavlov, E., Vaikuntanathan, V.: Byzantine agreement in the full-information model in $O(\log n)$ rounds. In: Proc. 38th Annual ACM Symposium on Theory of Computing. p. 179–186. STOC'06, ACM (2006)
10. Beraldi, R., Mattia, G.: Power of random choices made efficient for fog computing. IEEE Transactions on Cloud Computing **10**(02), 1130–1141 (2022)
11. Berenbrink, P., Czumaj, A., Steger, A., Vöcking, B.: Balanced allocations: The heavily loaded case. SIAM Journal on Computing **35**(6), 1350–1385 (2006)
12. Bramson, M., Lu, Y., Prabhakar, B.: Asymptotic independence of queues under randomized load balancing. Queueing Syst. Theory Appl. **71**(3), 247–292 (2012)

13. Busch, C., Kowalski, D.R.: Byzantine-resilient population protocols. CoRR **abs/2105.07123** (2021)
14. Chrobak, M., Gasieniec, L., Kowalski, D.R.: The wake-up problem in multihop radio networks. SIAM J. Comput. **36**(5), 1453–1471 (2007)
15. Cole, R., Maggs, B.M., Meyer auf der Heide, F., Mitzenmacher, M., Richa, A.W., Schröder, K., Sitaraman, R.K., Vöcking, B.: Randomized protocols for low-congestion circuit routing in multistage interconnection networks. In: Proc., 30th Annual ACM Symposium on Theory of Computing. p. 378–388. STOC'98 (1998)
16. Cooper, C., Elsässer, R., Radzik, T.: The power of two choices in distributed voting. In: Esparza, J., Fraigniaud, P., Husfeldt, T., Koutsoupias, E. (eds.) Automata, Languages, and Programming. pp. 435–446. Springer Berlin Heidelberg (2014)
17. Gonnet, G.H.: Expected length of the longest probe sequence in hash code searching. J. ACM **28**(2), 289–304 (apr 1981)
18. Guerraoui, R., Ruppert, E.: Names trump malice: Tiny mobile agents can tolerate byzantine failures. In: Automata, Languages and Programming, 36th Internatilonal Colloquium, ICALP 2009. LNCS, vol. 5556, pp. 484–495. Springer (2009)
19. Hagerup, T., Rüb, C.: A guided tour of chernoff bounds. Information Processing Letters **33**(6), 305–308 (1990)
20. Johnson, N.L., Kotz, S.: Urn models and their application : an approach to modern discrete probability theory. International Statistical Review **46**, 319 (1978)
21. Kenthapadi, K., Panigrahy, R.: Balanced allocation on graphs. In: ACM-SIAM Symposium on Discrete Algorithms (SODA). p. 434–443 (2006)
22. Lamport, L., Shostak, R., Pease, M.: The byzantine generals problem. ACM Transactions on Programming Languages and Systems pp. 382–401 (July 1982)
23. Los, D., Sauerwald, T.: Balanced allocations with the choice of noise. In: Proc., ACM Sym. on Principles of Distributed Computing (PODC'22). p. 164–175 (2022)
24. Lynch, N.A.: Distributed Algorithms. Morgan Kaufmann, 1st edn. (1996)
25. Park, C.J.: Random allocations (valentin f. kolchin, boris a. sevast'yanov and vladimir p. chistyakov). Siam Review **22**, 104–104 (1980)
26. Peres, Y., Talwar, K., Wieder, U.: Graphical balanced allocations and the $1+\beta$-choice process. Random Struct. Algorithms **47**(4), 760–775 (Dec 2015)
27. Richa, A., Mitzenmacher, M., Sitaraman, R.: The power of two random choices: A survey of techniques and results. Handbook of Randomized Computing (2001)
28. Vöcking, B.: How asymmetry helps load balancing. In: Proc. of the 40th Annual IEEE Symposium on Foundations of Computer Science. p. 131. FOCS'99 (1999)