Alcom-FT Technical Report Series
ALCOMFT-TR-03-72

Modeling Temporal Locality in Paging: A tighter Analysis of LRU and FWF *

Luca Becchetti †

† Università di Roma "La Sapienza", Italy, `becchett@dis.uniroma1.it`.

Abstract

In this paper, we extend the diffuse adversary approach for paging initially proposed in [11], with the aim of modelling locality of reference. We consider two very general families of distributions, parametrized by their expected values and standard deviations.

We prove that LRU's competitive ratio rapidly becomes constant as the standard deviation becomes small, a result that is further improved when the distribution satisfies a concavity property.

In contrast, we prove that FWF performs poorly with respect to the distributions e propose. We think this result provides one theoretical explanation of a phenomenon that is very well known in practice, but that was so far not completely assessed by theory.

*Partially supported by the EU project AMORE grant HPRN-CT-1999-00104, IST Programme of the EU under contract number IST-1999-14186 (ALCOM-FT), IST-2000-14084 (APPOL), IST-2001-33555 (COSIN) and by the MIUR Project "Resource Allocation in Real Networks".

1 Introduction

Paging is the problem of managing a two-level memory consisting of a first level fast memory or *cache* and a second level slower memory, both divided into pages of equal, fixed size. We assume that the slow memory contains a universe $\mathcal{P}$ of pages that can be requested, while the cache can host a fixed number k of pages. Normally, $|\mathcal{P}| \gg k$.

The input to a paging algorithm is a sequence of on-line requests $\sigma = \{\sigma(1), \ldots, \sigma(n)\}$, each specifying a page to access. The i-th request is said to have *index* i. If the page is in main memory the access operation costs nothing. Otherwise, a page fault occurs: a page in cache has to be evicted, in order to make room for the requested page, which has to be brought into cache before being accessed. In this case, the cost of the operation is 1. Each page has to be served before the next request is presented. The goal is minimizing the total cost, i.e. the number of page faults. In the following, if $\overline{\sigma}$ and $\hat{\sigma}$ are request sequences over $\mathcal{P}$, $\sigma = \overline{\sigma}\hat{\sigma}$ denotes the new request sequence obtained when all requests in $\overline{\sigma}$ are presented before every request in $\hat{\sigma}$. $\overline{\sigma}$ and $\hat{\sigma}$ are said to be, respectively, a *prefix* and a *postfix* of σ. We use $|\sigma|$ to denote the length of σ. Given σ, by $\sigma(i)$'s prefix we mean the subsequence $\sigma(1) \cdots \sigma(i-1)$.

The output of a paging algorithm corresponding to request sequence σ is the sequence $\mathcal{C}(\sigma_1), \ldots, \mathcal{C}(\sigma_{|\sigma|})$ of cache states, with $\sigma_1 = \sigma(1)$ and $\sigma_i = \sigma_{i-1}\sigma(i)$ for $i > 1$.

Paging algorithms. The core of a paging algorithm is the strategy according to which a page to evict is chosen when a fault occurs. Obviously, a paging algorithm should avoid evicting pages that are going to be requested in the near future. In this paper we want to analyze the performance of two well known heuristics for page eviction, namely LRU (Least Recently Used) and FWF (Flush When Full). The former is maybe the best known algorithm for virtual memory management in operating systems [15] and has found application in web caching as well. The latter has mostly theoretical interest, as explained below.

- **Least Recently Used (LRU):** When a page fault occurs, replace the page whose most recent request was earliest.
- **Flush When Full (FWF):** Assume the cache is initially empty. Three cases may occur by a request: i) the page requested is in cache: in this case the request is served at no extra cost; ii) a page fault occurs and the cache is not full: in this case the page is brought into cache, no page is evicted; iii) a page fault occurs and the cache is full: in this case, *all* pages in the cache are evicted, i.e. the cache is flushed.

Competitive analysis of paging algorithms. Sleator and Tarjan [14] introduced competitive analysis and were the first to show that $FIFO$ and LRU are k-competitive.

Surprisingly, FWF also has the same competitive ratio [2]. In contrast, LRU and FWF behave very differently in practice [17]: LRU is almost optimal on most instances, whereas FWF's performance is often poor. This discrepancy motivates the (essentially theoretical) interest of FWF: In a sense, the behaviour of this algorithm is an evidence for the inaccuracy of traditional competitive analysis in assessing the performance of paging algorithms. The main cause for this is the inadequacy of traditional competitive analysis to model a fundamental phenomenon that makes some paging algorithms work in practice: locality. *Locality of reference* [15] has a great impact on the actual behaviour of paging algorithms. Roughly, locality means that pages that have been accessed recently in the past (temporal locality) or whose position in the address space is close to pages that have been accessed recently (spatial locality) are likely to be accessed in the next future.

Considerable efforts have been devoted to closing the gap between theory and practice of paging algorithms, with particular attention to locality of reference. The early work of Franaszek and Wagner [8] studies the behaviour of several algorithms, including FIFO and LRU using probabilistic analysis. There analysis, however, does not model locality. Fiat et al. [5] explored the role of randomization, proposing a randomized paging algorithm that is H_k-competitive.

Borodin et al. [3] were the first to propose extensions to standard competitive analysis. In [3], the authors assume that pages are vertices of an underlying graph, called *access graph*. Locality is modelled by constraining request sequences to correspond to simple walks in the access graph. The same model was used and extended in subsequent works [9, 6, 12]. These works show that the competitiveness of LRU is never worse than FIFO in the access graph model (see in particular [12]), while [9] proposes an algorithm FAR that asymptotically achieves the best performance among on-line algorithms. This latter result was extended in [7] to multi-pointer walks in the access graph.

The work of [10] is similar in spirit, although the approach is probabilistic: the authors assume that the request sequence is generated by a stochastic process resulting from a Markov chain.

In [16], Torng models locality by considering the relationship between the length of a request sequence and the number of distinct pages that are actually requested. He further considers the cost model in which access to a page in fast memory costs 1, while a page fault costs $p > 1$. He proves that all marking algorithms achieve the best possible competitive ratio in this model.

Young [17, 18] introduces the notion of loose competitiveness. Roughly speaking, a paging algorithm is loosely $c(k)$-competitive if, for any request sequence, for most values of the cache size k, the cost of the algorithm is within $c(k)$ times the optimum.

Albers et al. [1] model locality using the working set concept proposed by Denning [4]. In particular, they assume that the maximum/average number of distinct page requests over a sequence of length n is a concave function in n. The cost function they consider is fault rate which, for a sequence σ of length n, is defined as the ratio between the number of page faults incurred by the algorithm and n. They then study the performance of several algorithms, including marking algorithms, FIFO and LFD, in terms of this metric. They prove that, among on-line algorithms, LRU achieves best in the first model, while both LRU and FIFO achieve best, among on-line algorithms, in the second model.

Koutsoupias and Papadimitriou introduce the notion of *diffuse adversary*. In this model, the adversary is constrained to generate the input sequence according to a given family of probability distributions. In [11] the authors make the following assumption on the probability distribution of the request sequence: at any stage, the probability that some page p is requested, conditioned to any request sequence prior to p, is bounded by some $\epsilon > 0$. The authors prove that, under this family of distributions, LRU achieves the best possible competitive ratio among on-line algorithms. These results where extended by Young [19, 20], who bounded the competitive ratio of marking algorithms against the diffuse adversary described above. In particular, he proved that the competitive ratio fastly becomes constant as $\epsilon = O(1/k)$, is $\Theta(\log k)$ when $\epsilon = 1/k$ and rapidly becomes $\Omega(k)$ as $\epsilon = \Omega(1/k)$.

Contribution of this paper. In this paper, we extend the diffuse adversary approach for paging, proposing an adversary that, differently from [11], attempts to model locality of reference, in particular temporal locality. Alike [11], our adversary is also defined in terms of the probability that a page p is requested, conditioned to some request sequence prior to p. Differently from [11] however, this probability is a non-increasing function of p's age, i.e. the more recently a page was requested, the higher the probability the page is requested again. We consider two very general families of distributions, parametrized by their expected values and standard deviations. We prove that LRU's competitive ratio rapidly becomes constant as the standard deviation becomes small, a result that is further improved when the distribution satisfies a concavity property defined further in this paper.

In contrast, we prove that FWF performs very poorly with respect to the distributions e propose. This result in our opinion provides one theoretical validation of the model we propose and an explanation of a phenomenon that is very well known in practice, but that was so far not fully explained. Observe that the results for LRU require specific properties of this algorithm and do not necessarily apply to other marking algorithms.

We also point out that the distributions we consider may include ones for which the diffuse adversary considered in [11] and [19, 20] brings to a poor upper bound for LRU.

Roadmap. In Section 2 we propose our model and discuss some preliminaries. In Section 3 we provide a diffuse adversary analysis of LRU, where the distributions allowed to the adversary are used to model temporal locality. We prove that LRU is essentially constant competitive with respect to a wide class of distributions. We also prove that, under the same diffuse adversary, the performance of FWF is extremely poor. So, the adversary we propose seems to model some crucial aspects of the problem, that were not completely captured before.

For the sake of brevity and readibility, many proofs have been moved to the appendix.

2 Model and preliminaries

We can partition each request sequence into *phases* in the usual way [2]. In particular, phases are defined inductively as follows: i) phase 1 starts with the first request and completes when k *distinct* pages have been requested; ii) assuming phase $i-1$, phase i starts with the $k+1$-th distinct request after the start of phase $i-1$. We point out that the last phase of a request sequence might contain less than k distinct requests. However, it has to contain at least one, by the definition of phase.

We say that a page is *marked* in a phase, if it was already requested during the phase. Otherwise the page is said *unmarked*. A *marking* algorithm never evicts a marked page when a page fault occurs. Observe that both LRU and FWF are marking algorithms.

Diffuse adversary. The *diffuse adversary* model was proposed by Koutsoupias and Papadimitriou [11]. It removes the overly pessimistic assumption that we know nothing about the distribution according to which the input is generated and assumes that the actual distribution D according to which the input is generated is member of a known family Δ of possible distributions. We say that algorithm A is c-competitive against the Δ-diffuse adversary if, for every distribution $D \in \Delta$ of the input sequence,

$$\mathbf{E}_D[A(\sigma)] \leq c\mathbf{E}_D[\mathrm{OPT}(\sigma)] + b,$$

where σ is generated according to D and b is a non-negative constant. The *competitive ratio* of A against the Δ-diffuse adversary is the minimum c such that the condition above holds. In the sequel, we drop D from the expression of the expectation when clear from context.

The request sequence is a stochastic process. Considering without loss of generality input sequences of length n, we define the following random variables:

- If $i \leq n$, $\Gamma(i)$ denotes the i-th page requested.
- For $n \geq l \geq i$, $\Gamma(i, l)$ denotes the subsequence of indices $i, \ldots, l$.

- If $i \leq n$, $\overline{\Gamma}(i)$ denotes the prefix of the i-th request.
- If $i \leq n$, $\hat{\Gamma}(i)$ denotes the postfix after the i-th request. Of course, $\hat{\Gamma}(i)$ is deterministically the empty string if $i = n$.

Similarly to [11] and [19], the family of distributions we consider is defined by specifying, for every page $p \in \mathcal{P}$ and for every possible prefix σ, with $|\sigma| = i - 1 < n$, the probability $\mathbf{P}[p \,|\, \sigma]$ that p is requested, provided σ is the prefix. More in detail, we assume that $\mathbf{P}[p \,|\, \sigma]$ belongs to some family Δ. The distributions we consider are intended to model temporal locality of page references. The general result that seems to emerge from our analysis is intuitive and well known in practice: *algorithms that keep the most likely pages in cache perform well.* We remark that this statement does not exclude different strategies from performing as.

In order to define the distributions we are interested in, we need the following definition:

Definition 1. Consider a monotone, non increasing function $f : I \rightarrow (0, 1)$, where $I = \{0, \ldots, N\}$ for some positive integer N. We say that f is **concave** in I if, for all $i = 1, \ldots, N-1$: $f(i-1) - f(i) \leq f(i) - f(i+1)$. f is **convex** in I if, for all $i = 1, \ldots, N-1 : f(i-1) - f(i) \geq f(i) - f(i+1)$.

We next describe distributions of the input sequence that try to model temporal locality, which is known to be the major cause of success of LRU and LRU-like heuristics in practice. To this aim, assume σ is the request sequence observed so far. Provided both pages p and q appear in request sequence σ, we say that p is *younger* than q if the last request for p in σ occurs after the last request for q, while we say that p is *older* than q otherwise. The age of any page p with respect to σ is then defined as follows: $\mathbf{age}(p, \sigma) = l$, if p is the l-th most recently accessed page, while $\mathbf{age}(p, \sigma) = \infty$ if p does not appear in σ. In order to describe the families of distributions we consider, we shall define, for every prefix sequence σ, a total order $\mathcal{R}(\sigma)$ on $\mathcal{P}$, that will in general depend on σ. Once a total oder is defined, there is a one-to-one correspondence between integers in $\{1, \ldots, |\mathcal{P}|\}$ and pages in $\mathcal{P}$. In the sequel, for any prefix sequence σ, we identify a page p with its corresponding element in $\{1, \ldots, |\mathcal{P}|\}$. Hence, for instance, $(\Gamma(i) = p \,|\, \overline{\Gamma}(i) = \sigma)$ denotes the event that the i-th request is for the p-th page in the total order $\mathcal{R}(\sigma)$. We can now give the following

Definition 2. Family $\mathcal{D}_1(\tau)$ (resp. $\mathcal{D}_2(\tau)$): Given any prefix σ, there is a total order $\mathcal{R}(\sigma)$ of pages in $\mathcal{P}$ that satisfies the following properties: (i) if $\mathbf{age}(p, \sigma) = l$, then p is the l-th element in $\mathcal{R}(\sigma)$; (ii) $\mathbf{P}[p \,|\, \sigma]$ is monotone, non-increasing and concave (resp. convex) in $\mathcal{R}(\sigma)$; (iii) $\mathbf{P}[p \,|\, \sigma]$ has expected value at most $k/2$ and standard deviation τ.

Some remarks are useful: i) $\mathcal{R}(\sigma)$ only has to obey property (i), while the distribution has to be such that also properties (ii) and (iii) are satisfied; ii) with this definition, if page p has already been requested in the past, the probability that p is requested again only depends on $\mathbf{age}(p, \sigma)$. So, for instance, if $\sigma_1 = \{p_7, p_3, p_4, p_4, p_4, p_2\}$ and $\sigma_2 = \{p_7, p_3, p_4, p_2\}$, then $\mathbf{P}[p_3 \,|\, \sigma_1] = \mathbf{P}[p_3 \,|\, \sigma_2]$.

There is one final issue as to how the stochastic process occurs: how is the i-th request generated? If $\mathbf{P}[p \,|\, \sigma]$ has to belong to either $\mathcal{D}_1(\tau)$ or $\mathcal{D}_2(\tau)$ for all $i \leq n$ then we speak of *concave* (resp. *convex*) distributions. Conversely, if $\mathbf{P}[p \,|\, \sigma]$ belongs to $\mathcal{D}_1(\tau) \cup \mathcal{D}_2(\tau)$ we speak of *mixed* distributions. In Section 3.1 we obtain a stronger result for concave distributions, while the result for convex ones is basically the same as for mixed.

3 Analysis of LRU

Considered any phase $\ell \geq 1$, we say that a page p is *new* in ℓ if it was not requested during phase $\ell - 1$. Otherwise we say that p is *old*. Considered any point in the ℓ-th phase (if any) of the request sequence, we say the page p is *critical*, if p is old but it is not currently in cache. Considered any prefix σ, we denote by $\mathcal{C}(\sigma)$, $\mathcal{N}(\sigma)$ and $\mathcal{O}(\sigma)$ respectively the sets of in cache, new and critical pages.

In the sequel, W^ℓ and N^ℓ are random variables that denote the number of requests to, respectively, critical and new pages during phase ℓ, while Z^ℓ denotes the overall number of requests to pages in cache during the phase. $W^\ell = N^\ell = Z^\ell = 0$ if no ℓ-th phase exists. Finally, D^ℓ denotes the number of distinct requests during the ℓ-th phase. Alike the previous variables, $D^\ell = 0$ if no ℓ-th phase exists. W^ℓ_σ, N^ℓ_σ, N^ℓ_σ and D^ℓ_σ respectively denote their deterministic values for a particular request sequence σ. Obviously, they are all equal to 0 when σ does not contain an ℓ-th phase. Furthermore, $D^\ell_\sigma = k$, unless ℓ is the last phase in σ and it is not complete. The following result is due to Young [19]:

Theorem 1 ([19]). *For every request sequence σ:*

$$\frac{1}{2}\sum_\ell N^\ell_\sigma \leq OPT(\sigma) \leq \sum_\ell N^\ell_\sigma.$$

The following facts are well known:

Fact 1. *If request sequence σ contains L phases then $OPT(\sigma) \geq L$. More precisely, OPT faults at least once per phase.*

Fact 2. *At any point of the execution, the k most recent pages are in the cache of LRU.*

The main goal of this section is to bound

$$\frac{\mathbf{E}[\mathrm{LRU}(\sigma)]}{\mathbf{E}[\mathrm{OPT}(\sigma)]},$$

when σ follows any probability distribution belonging to the families $\mathcal{D}_1(\tau)$ or $\mathcal{D}_2(\tau)$. Let F be the random variable that denotes the overall number of phases. We can write:

$$\frac{\mathbf{E}[\mathrm{LRU}(\sigma)]}{\mathbf{E}[\mathrm{OPT}(\sigma)]} \leq 2\frac{\mathbf{E}\left[\sum_{\ell=1}^F (W^\ell + N^\ell)\right]}{\mathbf{E}\left[\sum_{\ell=1}^F N^\ell\right]} = 2\frac{\mathbf{E}\left[\sum_{\ell=1}^F W^\ell\right]}{\mathbf{E}\left[\sum_{\ell=1}^F N^\ell\right]} + 2, \tag{1}$$

where the first inequality follows from Theorem 1. We continue with:

$$\mathbf{E}\left[\sum_{\ell=1}^F W^\ell\right] = \sum_{L\geq 1}\mathbf{E}\left[\sum_{\ell=1}^L W^\ell \,|\, F = L\right]\mathbf{P}[F = L]. \tag{2}$$

For the sake of the analysis, it is convenient to express Equation (2) in a different way, namely:

$$\begin{aligned}\sum_{L\geq 1}\mathbf{E}\left[\sum_{\ell=1}^L W^\ell \,|\, F = L\right]\mathbf{P}[F = L] &= \sum_{L\geq 1}\sum_{\ell=1}^L \mathbf{E}\left[W^\ell \,|\, F = L\right]\mathbf{P}[F = L] \\ = \sum_{\ell\geq 1}\sum_{L\geq \ell}\mathbf{E}\left[W^\ell \,|\, F = L\right]\mathbf{P}[F = L].\end{aligned}$$

On the other hand:

$$\mathbf{E}\Big[W^\ell \,|\, F \geq \ell\Big] = \sum_{L \geq \ell} \mathbf{E}\Big[W^\ell \,|\, F = L\Big]\, \mathbf{P}[F = L \,|\, F \geq \ell] = \sum_{L \geq \ell} \mathbf{E}\Big[W^\ell \,|\, F = L\Big] \frac{\mathbf{P}[F = L]}{\mathbf{P}[F \geq \ell]},$$

which implies that

$$\sum_{L \geq \ell} \mathbf{E}\Big[W^\ell \,|\, F = L\Big]\, \mathbf{P}[F = L] = \mathbf{E}\Big[W^\ell \,|\, F \geq \ell\Big]\, \mathbf{P}[F \geq \ell],$$

and therefore:

$$\mathbf{E}\left[\sum_{\ell=1}^{F} W^\ell\right] = \sum_{\ell \geq 1} \mathbf{E}\Big[W^\ell \,|\, F \geq \ell\Big]\, \mathbf{P}[F \geq \ell]. \tag{3}$$

Analogously:

$$\mathbf{E}\left[\sum_{\ell=1}^{F} N^\ell\right] = \sum_{\ell \geq 1} \mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big]\, \mathbf{P}[F \geq \ell]. \tag{4}$$

Therefore:

$$\frac{\mathbf{E}[\mathrm{LRU}(\sigma)]}{\mathbf{E}[\mathrm{OPT}(\sigma)]} \leq 2\frac{\sum_{\ell \geq 1} \mathbf{E}\big[W^\ell \,|\, F \geq \ell\big]\, \mathbf{P}[F \geq \ell]}{\sum_{\ell \geq 1} \mathbf{E}[N^\ell \,|\, F \geq \ell]\, \mathbf{P}[F \geq \ell]} + 2. \tag{5}$$

We now introduce the following random variables, for all pairs (ℓ, j), where $j = 1, \ldots, k$: i) $X_j^\ell = 1$ if there is an ℓ-th phase, there is a distinct j-th request and this request is for a critical page, $X_j^\ell = 0$ otherwise (i.e. no ℓ-th phase exists or the phase is not complete and it does contain a j-th distinct request or it does, but this request is not for a critical page); ii) $Y_j^\ell = 1$ if the $j - th$ distinct request in phase ℓ is to a new page, $Y_j^\ell = 0$ otherwise; iii) finally, $Z_j^\ell = 1$ if the $j - th$ distinct request in phase ℓ is to a page in cache, $Z_j^\ell = 0$ otherwise. A consequence of the definitions above is that we have $\sum_{j=1}^{k} X_j^\ell = W^\ell$ and analogously for N^ℓ and Z^ℓ.

The next step of our proof concerns relating $\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell]$ to $\mathbf{P}[Y_j^\ell = 1 \,|\, F \geq \ell]$ and $\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell]$. Intuitively, we prove that i) if $\mathbf{P}[p \,|\, \sigma]$ is concave then the average number of requests to critical pages in the ℓ-th phase is comparable with the average number of requests to new pages in the phase; ii) if $\mathbf{P}[p \,|\, \sigma]$ is convex, then with high probability, most of the requests in the phase are for pages in cache, with the exception of a small number that rapidly decreases as $\mathbf{P}[p \,|\, \sigma]$ becomes concentrated around pages in cache. We point out that in both cases the proof requires the property of LRU, ensuring that the k most likely pages are in cache.

3.1 Concave distributions

In this section we provide an analysis of LRU for the case of concave distributions. We point out that, for concave distributions and provided $|\mathcal{P}| \geq 4k$, our result holds without any assumptions on the values of expectation and standard deviation. We start with the following Lemma:

Lemma 1. *If* $\mathbf{P}[p \mid \sigma]$ *belongs to family* $\mathcal{D}_1(\tau)$ *and* $|\mathcal{P}| \geq 4k$,

$$\mathbf{P}[X_j^\ell = 1 \mid F \geq \ell] \leq 2\mathbf{P}[Y_j^\ell = 1 \mid F \geq \ell].$$

Proof. Observe that $\mathbf{P}[X_j^\ell = 1 \mid F \geq \ell] = 0$ if $j = 1$, since the first distinct request in the phase cannot be to a critical page. For the rest of this proof, denote by $\mathcal{S}(\ell, t)$, $t \geq 1$, the set of prefixes σ such that σ contains ℓ phases and t distinct requests during the ℓ-th phase. For $j \geq 2$ we can write:

$$\mathbf{P}[X_j^\ell = 1 \mid F \geq \ell] = \sum_{\sigma \in \mathcal{S}(\ell, j-1)} \mathbf{P}[\overline{\Gamma}(|\sigma|+1) = \sigma \cap \Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \mid F \geq \ell]$$
$$= \frac{1}{\mathbf{P}[F \geq \ell]} \sum_{\sigma \in \mathcal{S}(\ell, j-1)} \mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \mid \overline{\Gamma}(|\sigma|+1) = \sigma] \cdot \mathbf{P}[\overline{\Gamma}(|\sigma|+1) = \sigma],$$

where the second equality follows since $(\overline{\Gamma}(|\sigma|+1) = \sigma \in \mathcal{S}(\ell, j-1))$ implies $(F \geq \ell)$. Similarly:

$$\mathbf{P}[Y_j^\ell = 1 \mid F \geq \ell] = \sum_{\sigma \in \mathcal{S}(\ell, j-1)} \mathbf{P}[\overline{\Gamma}(|\sigma|+1) = \sigma \cap \Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \mid F \geq \ell]$$
$$= \frac{1}{\mathbf{P}[F \geq \ell]} \sum_{\sigma \in \mathcal{S}(\ell, j-1)} \mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \mid \overline{\Gamma}(|\sigma|+1) = \sigma] \cdot \mathbf{P}[\overline{\Gamma}(|\sigma|+1) = \sigma].$$

The following Lemma holds:

Lemma 2. *If* $\mathbf{P}[p \mid \sigma]$ *belongs to family* $\mathcal{D}_1(\tau)$ *and* $|\mathcal{P}| \geq 4k$,

$$\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \mid \overline{\Gamma}(|\sigma|+1) = \sigma] \leq 2\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \mid \overline{\Gamma}(|\sigma|+1) = \sigma].$$

Proof. See appendix. □

From Lemma 2 the proof immediately follows. □

Corollary 1. *If* $\mathbf{P}[p \mid \sigma]$ *belongs to family* $\mathcal{D}_1(\tau)$ *and* $|\mathcal{P}| \geq 4k$,

$$\frac{\mathbf{E}[W^\ell \mid F \geq \ell]}{\mathbf{E}[N^\ell \mid F \geq \ell]} \leq 2.$$

Proof. The proof immediately follows from Lemma 1. □

Corollary 1 and Equation (5) immediately imply the first, following Theorem:

Theorem 2. *If* $\mathbf{P}[p \mid \sigma]$ *belongs to family* $\mathcal{D}_1(\tau)$ *and* $|\mathcal{P}| \geq 4k$,

$$\frac{\mathbf{E}[LRU(\sigma)]}{\mathbf{E}[OPT(\sigma)]} \leq 6.$$

Observe that Theorem 2 does not necessarily hold for other marking algorithms, since the proof of Lemma 2 requires properties (i) and (ii) of Definition 2 to hold.

3.2 Mixed distributions

We now turn to the case in which $\mathbf{P}[p \mid \sigma]$ belongs to family $\mathcal{D}_1(\tau) \cup \mathcal{D}_2(\tau)$, the analysis for the convex case being substantially the same. In this case, the analysis turns out to be more complex, since we cannot always bound the expected number of requests to critical pages with the expected number of requests to new pages. In this case, simply put, two phenomena are possible: either the expected numbers of requests to new and critical pages are related, as in the previous case, or, with high probability, most of the requests are for pages in cache. The following lemma will be useful:

Lemma 3. *If $\mathbf{P}[p \mid \sigma]$ belongs to $\mathcal{D}_1(\tau) \cup \mathcal{D}_2(\tau)$ then*

$$\mathbf{P}[\Gamma(i) \notin \mathcal{C}(\sigma) \mid \overline{\Gamma}(i) = \sigma] \leq \alpha(k,\tau) = \left(\frac{2\tau}{k}\right)^2 .$$

Proof. As remarked above, there is a total order over pages, with the first k pages of this order in cache. So, defined by this order, there is a one-to-one correspondence between elements of $\mathcal{P}$ and elements in $\{1, 2, \ldots, |\mathcal{P}|\}$, with i corresponding to the i-th most recently requested page in cache[1]. $\mathbf{P}[p \mid \sigma]$ may then be viewed as a probability distribution over $\{1, 2, \ldots, |\mathcal{P}|\}$, with expected value at most $k/2$ and standard deviation τ. The probability that a page not in cache is requested is then the probability that an integer larger than k is picked. Applying Chebyshev's inequality [13] the claim follows. □

We next bound the number of new pages that are requested with high probability, namely, we prove that, conditioned to $(F \geq \ell)$, $(N^\ell \leq x)$ with high probability, where x is suitably small. More precisely, we prove the following Lemma:

Lemma 4. *If $\mathbf{P}[p \mid \sigma]$ belongs to family $\mathcal{D}_1(\tau) \cup \mathcal{D}_2(\tau)$ and*

$$x = \left\lceil \frac{\log k + \log \frac{1}{1-\alpha(k,\tau)}}{\log \frac{1}{\alpha(k,\tau)}} \right\rceil ,$$

then,

$$\mathbf{P}[N^\ell > x \mid F \geq \ell] \leq \frac{1}{k}.$$

Proof. For the rest of this proof, denote by $\overline{P}(t)$ the set of prefixes the contain t complete phases and, for any prefix $\overline{\sigma} \in \overline{P}(\ell - 1)$, denote by $\hat{P}(\overline{\sigma}, y)$ the set of subsequences $\hat{\sigma}$ such that, given $\overline{\sigma}$ as a prefix, $\hat{\sigma}$ contains exactly y requests for new pages and $|\overline{\sigma}\hat{\sigma}| \leq n$. We define the following event, for brevity: $\mathcal{F}(\sigma) = (\Gamma(1, |\sigma|) = \sigma)$. Given these definitions we can write:

$$\mathbf{P}[N^\ell = y \mid F \geq \ell] = \sum_{\overline{\sigma} \in \overline{P}(\ell-1)} \sum_{\hat{\sigma} \in \hat{P}(\overline{\sigma}, y)} \mathbf{P}[\mathcal{F}(\overline{\sigma}\hat{\sigma}) \mid F \geq \ell]$$

$$= \sum_{\overline{\sigma} \in \overline{P}(\ell-1)} \sum_{\hat{\sigma} \in \hat{P}(\overline{\sigma}, y)} \mathbf{P}[\Gamma(|\overline{\sigma}| + 1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \mid \mathcal{F}(\overline{\sigma}) \cap F \geq \ell] \, \mathbf{P}[\mathcal{F}(\overline{\sigma}) \mid F \geq \ell]$$

Now,

$$\mathbf{P}[\Gamma(|\overline{\sigma}| + 1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \mid \mathcal{F}(\overline{\sigma}) \cap F \geq \ell] = \frac{\mathbf{P}[\Gamma(|\overline{\sigma}| + 1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \mid \mathcal{F}(\overline{\sigma})]}{\mathbf{P}[F \geq \ell \mid \mathcal{F}(\overline{\sigma})]}$$

[1] Observe that, in order to prove this lemma, we need no assumption on the order of pages not in cache, any order is valid.

$$= \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \,|\, \mathcal{F}(\overline{\sigma})]}{\mathbf{P}[D^\ell \geq 1 \,|\, \mathcal{F}(\overline{\sigma})]},$$

where the first equality follows since, for $\overline{\sigma} \in \overline{P}(\ell-1)$ and $\hat{\sigma} \in \hat{P}(\overline{\sigma}, y)$, the event $(\Gamma(1, |\overline{\sigma}\hat{\sigma}|) = \overline{\sigma}\hat{\sigma})$ implies $(F \geq \ell)$, while the second follows since $(D^\ell \geq 1)$ implies $(F \geq \ell)$ and viceversa.

For any pair $(\overline{\sigma}, \hat{\sigma})$, we denote by $\mathcal{I}(\overline{\sigma}, \hat{\sigma})$ the subset of $\{1, \ldots, |\hat{\sigma}|\}$ such that, provided $\overline{\sigma}$ is a prefix of $\hat{\sigma}$, $\hat{\sigma}(i)$ is a request to a new page if and only $i \in \mathcal{I}(\overline{\sigma}, \hat{\sigma})$. We have:

$$\begin{aligned}
&\mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \,|\, \mathcal{F}(\overline{\sigma})] \\
&= \Pi_{i=1}^{|\hat{\sigma}|} \mathbf{P}[\Gamma(|\overline{\sigma}|+i) = \hat{\sigma}(i) \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, \Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1}] \\
&\cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1} \,|\, \mathcal{F}(\overline{\sigma})],
\end{aligned}$$

where the second equality follows by a standard manipulation. Set $f = \min\{i : i \in \mathcal{I}(\overline{\sigma}, \hat{\sigma})\}$. So, $|\overline{\sigma}|+f$ is the index of the first request to a new page in the ℓ-th phase of the overall sequence $\overline{\sigma}\hat{\sigma}$. We proceed as follows:

$$\begin{aligned}
&\Pi_{i=1}^{|\hat{\sigma}|} \mathbf{P}[\Gamma(|\overline{\sigma}|+i) = \hat{\sigma}(i) \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, \Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1}] \\
&\cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1} \,|\, \mathcal{F}(\overline{\sigma})] \\
&\leq \alpha(k,\tau)^{y-1} \mathbf{P}[\Gamma(|\overline{\sigma}|+f) = \sigma(f) \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, \Gamma(|\overline{\sigma}|+1, f+i-1) = \hat{\sigma}_{f-1}] \\
&\cdot \Pi_{i \notin \mathcal{I}(\overline{\sigma}, \hat{\sigma})} \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i) = \hat{\sigma}_i \,|\, \mathcal{F}(\overline{\sigma})],
\end{aligned}$$

where the inequality follows since, provided $\overline{\sigma}$ is the prefix, there are exactly y requests for new pages during subsequence $\hat{\sigma}$, each with probability at most $\alpha(k, \tau)$ and, for all $i = 1, \ldots, |\hat{\sigma}|$:

$$\begin{aligned}
&\mathbf{P}[\Gamma(|\overline{\sigma}|+i) = \hat{\sigma}(i) \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, \Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1}] \\
&\cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i-1) = \hat{\sigma}_{i-1} \,|\, \mathcal{F}(\overline{\sigma})] \;=\; \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i) = \hat{\sigma}_i \,|\, \mathcal{F}(\overline{\sigma})].
\end{aligned}$$

Now, set $l(\overline{\sigma}, \hat{\sigma}) = \max\{i : i \notin \mathcal{I}(\overline{\sigma}, \hat{\sigma})\}$. We have:

$$\Pi_{i \notin \mathcal{I}(\overline{\sigma}, \hat{\sigma})} \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i) = \hat{\sigma}_i \,|\, \mathcal{F}(\overline{\sigma})] \;\leq\; \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+l(\overline{\sigma}, \hat{\sigma})) = \hat{\sigma}_{l(\overline{\sigma}, \hat{\sigma})} \,|\, \mathcal{F}(\overline{\sigma})].$$

This brings:

$$\begin{aligned}
&\mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}\hat{\sigma}|) = \hat{\sigma} \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, F \geq \ell] \\
&\leq \alpha(k,\tau)^{y-1} \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+f) = \sigma(f) \,|\, \mathcal{F}(\overline{\sigma}) \,\cap\, \Gamma(|\overline{\sigma}|+1, f+i-1) = \hat{\sigma}_{f-1}]}{\mathbf{P}[D^\ell \geq 1 \,|\, \mathcal{F}(\overline{\sigma})]} \\
&\cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+l(\overline{\sigma}, \hat{\sigma})) = \hat{\sigma}_{l(\overline{\sigma}, \hat{\sigma})} \,|\, \mathcal{F}(\overline{\sigma})] \\
&= \alpha(k,\tau)^{y-1} \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+f) = \sigma(f) \,|\, \mathcal{F}(\overline{\sigma})]}{\mathbf{P}[D^\ell \geq 1 \,|\, \mathcal{F}(\overline{\sigma})]} \cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+l(\overline{\sigma}, \hat{\sigma})) = \hat{\sigma}_{l(\overline{\sigma}, \hat{\sigma})} \,|\, \mathcal{F}(\overline{\sigma})] \\
&\leq \alpha(k,\tau)^{y-1} \cdot \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+l(\overline{\sigma}, \hat{\sigma})) = \hat{\sigma}_{l(\overline{\sigma}, \hat{\sigma})} \,|\, \mathcal{F}(\overline{\sigma})].
\end{aligned}$$

The second equality follows since $(\Gamma(|\overline{\sigma}|+f) = \hat{\sigma}(f))$ only depends on the age of $\sigma(f)$ and, given $\mathcal{F}(\overline{\sigma})$, requests $\hat{\sigma}(1), \ldots \hat{\sigma}(f-1)$ are for pages in cache by the definition of f and thus they do not change the age of $\hat{\sigma}(f)$. The third inequality follows since $(\Gamma(|\overline{\sigma}|+f) = \sigma(f) \,|\, \mathcal{F}(\overline{\sigma}))$ implies $(D^\ell \geq 1 \,|\, \mathcal{F}(\overline{\sigma}))$. Now, define $m = \min\{l(\overline{\sigma}, \hat{\sigma}) : \hat{\sigma} \in \hat{P}(\overline{\sigma}, y)\}$. We have:

$$\sum_{\hat{\sigma} \in \hat{P}(\overline{\sigma}, y)} \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+l(\overline{\sigma}, \hat{\sigma})) = \hat{\sigma}_{l(\overline{\sigma}, \hat{\sigma})} \,|\, \mathcal{F}(\overline{\sigma})]$$

$$\leq \sum_{\forall \hat{\sigma}:|\hat{\sigma}|\geq m} \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+m) = \hat{\sigma}_m \,|\, \mathcal{F}(\overline{\sigma})] \,\leq 1,$$

where the first inequality follows since, for fixed $\hat{\sigma} \in \hat{P}(\overline{\sigma}, y)$ and for $i_1 < i_2 \leq l(\overline{\sigma}, \hat{\sigma})$, $\mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i_1) = \hat{\sigma}_{i_1} \,|\, \mathcal{F}(\overline{\sigma})] \,\geq \mathbf{P}[\Gamma(|\overline{\sigma}|+1, |\overline{\sigma}|+i_2) = \hat{\sigma}_{i_2} \,|\, \mathcal{F}(\overline{\sigma})]$. As a consequence:

$$\mathbf{P}[N^\ell = y \,|\, F \geq \ell] \,\leq \alpha(k,\tau)^{y-1} \sum_{\overline{\sigma}\in\overline{P}(\ell-1)} \mathbf{P}[\mathcal{F}(\overline{\sigma}) \,|\, F \geq \ell] \,\leq \alpha(k,\tau)^{y-1}.$$

Now:

$$\mathbf{P}[N^\ell > x \,|\, F \geq \ell] \,\leq \sum_{y=x+1}^{k} \alpha(k,\tau)^{y-1} = \frac{\alpha(k,\tau)^x - \alpha(k,\tau)^k}{1-\alpha(k,\tau)} < \frac{\alpha(k,\tau)^x}{1-\alpha(k,\tau)}.$$

Choosing x according to the hypotheses, the latter expression is at most 1/k. This completes the proof. □

We next condition to the event $(N^\ell \leq x)$, x, roughly put, being an upper bound to the number of new pages requested during phase ℓ with high probability. We can state the following Lemma:

Lemma 5. *If $\mathbf{P}[p \,|\, \sigma]$ belongs to $\mathcal{D}_1(\tau) \cup \mathcal{D}_2(\tau)$ and $|\mathcal{P}| \geq 4k$,*

$$\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell] \,\leq 2\mathbf{P}[Y_j^\ell = 1 \,|\, F \geq \ell] + \frac{4}{t^2}\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell] + \frac{1}{k},$$

for $j = 1, \ldots, 2x$, where $t = \left\lfloor \frac{k-j}{x} \right\rfloor$.

Proof. Define by $\mathcal{S}(\ell, t, x)$, $t \leq k-1$, the set of pairs $(\overline{\sigma}, \hat{\sigma})$ such that: i) $\overline{\sigma}$ contains ℓ phases and t distinct requests in phase ℓ and ii) for some $p \in \mathcal{P}$, request sequence $\overline{\sigma}p\hat{\sigma}$ contains at most x requests to new pages in phase ℓ. If $\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] \,\leq 2\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$ the lemma is proved, so assume $\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] > 2\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$. Observe that $\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x] = 0$ for $j = 1$, hence, in the sequel, we assume $2 \leq j \leq k - 2x$. We have:

$$\begin{aligned}
&\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x] \,= \\
&\sum_{\substack{(\overline{\sigma},\hat{\sigma})\in \\ \mathcal{S}(\ell,j-1,x)}} \mathbf{P}[\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}) \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma} \,|\, F \geq \ell \cap N^\ell \leq x] \\
&= \frac{\sum_{\substack{(\overline{\sigma},\hat{\sigma})\in \\ \mathcal{S}(\ell,j-1,x)}} \mathbf{P}[\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}) \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma}]}{\mathbf{P}[F \geq \ell \cap N^\ell \leq x]} \\
&= \sum_{\substack{(\overline{\sigma},\hat{\sigma})\in \\ \mathcal{S}(\ell,j-1,x)}} \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}) \,|\, \overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma}]}{\mathbf{P}[F \geq \ell \cap N^\ell \leq x]} \\
&\cdot \mathbf{P}[\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma}] \\
&= \sum_{\substack{(\overline{\sigma},\hat{\sigma})\in \\ \mathcal{S}(\ell,j-1,x)}} \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}) \,|\, \overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma}]}{\mathbf{P}[F \geq \ell \cap N^\ell \leq x]} \cdot \mathbf{P}[\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma}]\,,
\end{aligned}$$

where the second equality follows since, if $(\overline{\sigma}, \hat{\sigma}) \in \mathcal{S}(\ell, j-1, x)$ and $p \in \mathcal{O}(\overline{\sigma})$, then the request sequence $\overline{\sigma}\hat{\sigma}$ implies the event $(F \geq \ell \cap N^\ell \leq x)$, while the fourth inequality follows since $(\Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}))$ depends on $(\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma})$ but not on $(\hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma})$. Analogously, we have:

$$\begin{aligned}
&\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x] \\
&= \sum_{\substack{(\overline{\sigma},\hat{\sigma})\in \\ \mathcal{S}(\ell,j-1,x)}} \frac{\mathbf{P}[\Gamma(|\overline{\sigma}|+1) \in \mathcal{C}(\overline{\sigma}) \,|\, \overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma}]}{\mathbf{P}[F \geq \ell \cap N^\ell \leq x]} \cdot \mathbf{P}[\overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma} \cap \hat{\Gamma}(|\overline{\sigma}|+1) = \hat{\sigma}]\,,
\end{aligned}$$

We continue with the following Lemma:

Lemma 6. *If $\mathbf{P}[p \,|\, \sigma]$ belongs to family $\mathcal{D}_2(\tau)$ and $\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] > 2\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$ then*

$$\mathbf{P}[\Gamma(|\overline{\sigma}|+1) \in \mathcal{O}(\overline{\sigma}) \,|\, \overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma}] \;\leq \frac{4}{t^2}\mathbf{P}[\Gamma(|\overline{\sigma}|+1) \in \mathcal{C}(\overline{\sigma}) \,|\, \overline{\Gamma}(|\overline{\sigma}|+1) = \overline{\sigma}]\,,$$

where $t = \left\lfloor \frac{k-j}{x} \right\rfloor$.

Proof. See appendix. □

Lemma 6 implies that

$$\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x] \;\leq \frac{4}{t^2}\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x]\,.$$

So, we have:

$$\begin{aligned}
&\mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell] \\
&= \mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x]\,\mathbf{P}[N^\ell \leq x \,|\, F \geq \ell] \\
&+ \mathbf{P}[X_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell > x]\,\mathbf{P}[N^\ell > x \,|\, F \geq \ell] \\
&\leq \frac{4}{t^2}\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell \cap N^\ell \leq x]\,\mathbf{P}[N^\ell \leq x \,|\, F \geq \ell] + \frac{1}{k} \\
&\leq \frac{4}{t^2}\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell] + \frac{1}{k},
\end{aligned}$$

where the second inequality follows from Lemma 6. □

Recalling the espression of t we can now write:

$$\begin{aligned}
&\mathbf{E}\Big[W^\ell \,|\, F \geq \ell\Big] \leq 2\mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big] + 16x^2 \sum_{j=1}^{k-2x} \frac{1}{(k-j)^2}\mathbf{P}[Z_j^\ell = 1 \,|\, F \geq \ell] + 2x \\
&\leq 2\mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big] + 16x^2 \sum_{j=1}^{k-2x} \frac{1}{(k-j)^2} + 2x \leq 2\mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big] + 16x^2 \int_{2x}^{k-1} \frac{1}{j^2}dj + 2x \\
&< 2\mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big] + 8x + 2x < 2\mathbf{E}\Big[N^\ell \,|\, F \geq \ell\Big] + 10x,
\end{aligned}$$

where the third inequality follows from simple manipulations. Finally, we have the following Theorem:

Theorem 3. *If* $\mathbf{P}[p\,|\,\sigma]$ *belongs to family* $\mathcal{D}_1(\tau)\cup\mathcal{D}_2(\tau)$ *and* $|\mathcal{P}|\geq 4k$,

$$\frac{\mathbf{E}[LRU(\sigma)]}{\mathbf{E}[OPT(\sigma)]}\leq 6+10\left\lceil\frac{\log k+\log\frac{1}{1-\alpha(k,\tau)}}{\log\frac{1}{\alpha(k,\tau)}}\right\rceil.$$

Proof. If OPT^ℓ is a random variable that gives the number of page faults for OPT during the ℓ-th phase, if any, we have:

$$\begin{aligned}\mathbf{E}[\mathrm{LRU}(\sigma)] &\leq \sum_{\ell\geq 1}(3\mathbf{E}\left[N^\ell\,|\,F\geq\ell\right]+10x)\mathbf{P}[F\geq\ell]\\ &\leq \sum_{\ell\geq 1}(6+10x)\mathbf{E}\left[\mathrm{OPT}^\ell\,|\,F\geq\ell\right]\mathbf{P}[F\geq\ell],\end{aligned}$$

where the second inequality follows from Theorem 1 and from Fact 1. Following the same lines as in the proof of Theorem 2 the claim follows. □

Alike Theorem 2, Theorem 3 does not necessarily hold for other marking algorithms, since the proof of Lemma 6 requires properties (i) and (ii) of Definition 2 to hold.

3.3 A lower bound for FWF

We prove that, under mixed, concave or convex distributions, FWF is basically $\Omega\,(k)$-competitive. This behaviour is in line with practical experience (see e.g. [17, 2]). Together with the results of Section 3, this seems to us one theoretical validation of the great difference observed between the behaviour of FWF and other marking algorithms (in particuar LRU).

Using the upper bound to the optimal cost of Theorem 1 and proceeding the same way as in Section 3, we have:

$$\frac{\mathbf{E}[\mathrm{FWF}(\sigma)]}{\mathbf{E}[\mathrm{OPT}(\sigma)]}\geq\frac{\sum_{\ell\geq 1}\mathbf{E}\left[W^\ell\,|\,F\geq\ell\right]\mathbf{P}[F\geq\ell]}{\sum_{\ell\geq 1}\mathbf{E}[N^\ell\,|\,F\geq\ell]\,\mathbf{P}[F\geq\ell]}+1,$$

where W^ℓ denotes the number of page faults on old pages of FWF.

We assume request sequences of arbitrary length, so that $\mathbf{E}[\mathrm{OPT}(\sigma)]$ can be arbitrarily large. Observe that, for each phase ℓ, W^ℓ is k, since the first request of the phase causes a page fault that FWF serves after evicting *all* pages currently in cache. As a consequence, the performance of FWF basically depends on the number of requests for new pages in each phase (i.e. the larger this number, the better the performance of the algorithm with respect to OPT). We have:

$$\begin{aligned}\mathbf{E}\left[W^\ell\,|\,F\geq\ell\right] &= \mathbf{E}\left[W^\ell\,|\,F\geq\ell\cap N^\ell\leq x\right]\mathbf{P}[N^\ell\leq x\,|\,F\geq\ell]\\ &+ \mathbf{E}\left[W^\ell\,|\,F\geq\ell\cap N^\ell> x\right]\mathbf{P}[N^\ell> x\,|\,F\geq\ell]\\ &\geq (k-x)\left(1-\frac{1}{k}\right),\end{aligned}$$

where the inequality follows from Lemma 4.

Proceeding in the same way, we have $\mathbf{E}\left[N^\ell\,|\,F\geq\ell\right]\leq x+1$ (since $N^\ell\leq k$ deterministically). Hence:

$$\frac{\mathbf{E}[\mathrm{FWF}(\sigma)]}{\mathbf{E}[\mathrm{OPT}(\sigma)]}\geq\frac{(k-x)\left(1-\frac{1}{k}\right)}{x+1}+1.$$

When the distribution becomes sufficiently concentrated around pages in cache (e.g. $\tau = O(\sqrt{k})$) the competitive ratio is $\Omega(k)$.

Acknowledgments. The author wishes to thank his colleagues and friends Stefano Leonardi and Alberto Marchetti-Spaccamela for their suggestions and support.

References

[1] Su. Albers, L. M. Favrholdt, and O. Giel. On paging with locality of reference. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC-02)*, pages 258–267, 2002.

[2] A. Borodin and R. El-Yaniv. *Online Computation and Competitive Analysis.* Cambridge University Press, 1998.

[3] A. Borodin, S. Irani, P. Raghavan, and B. Schieber. Competitive paging with locality of reference (preliminary version). In *Proceedings of the Twenty Third Annual ACM Symposium on Theory of Computing*, pages 249–259, 1991.

[4] P. J. Denning. The working set model for program behaviour. *Communications of the ACM*, 11(5):323–333, 1968.

[5] A. Fiat, R. Karp, M. Luby, L. McGeoch, D. Sleator, and N. E. Young. Competitive paging algorithms. 2002.

[6] A. Fiat and M. Mendel. Truly online paging with locality of reference. In *38th IEEE Annual Symposium on Foundations of Computer Science*, pages 326–335, 1997.

[7] Amos Fiat and Anna R. Karlin. Randomized and multipointer paging with locality of reference. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 626–634, 1995.

[8] P. A. Franaszek and T. J. Wagner. Some distribution-free aspects of paging performance. *Journal of the ACM*, 21:31–39, 1974.

[9] S. Irani, A. R. Karlin, and S. Phillips. Strongly competitive algorithms for paging with locality of reference. *SIAM Journal on Computing*, 25(3):477–497, 1996.

[10] A. R. Karlin, S. J. Phillips, and P. Raghavan. Markov paging (extended abstract). In *33rd Annual Symposium on Foundations of Computer Science*, pages 208–217, 1992.

[11] E. Koutsoupias and C. H. Papadimitriou. Beyond competitive analysis. In *Proceedings: 35th Annual Symposium on Foundations of Computer Science, November 20–22, 1994, Santa Fe, New Mexico*, pages 394–400, 1994.

[12] M.Chrobak and J. Noga. LRU is better than FIFO. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-98)*, pages 78–81, 1998.

[13] S. M. Ross. *Probability Models for Computer Science.* Adademic Press, 2002.

[14] D. Sleator and R. E. Tarjan. Amortized Efficiency of List Update and Paging Rules. *Communications of the ACM*, 28:202–208, 1985.

[15] Andrew S. Tanenbaum. *Modern Operating Systems.* Prentice Hall, 1992.

[16] E. Torng. A unified analysis of paging and caching. In *36th Annual Symposium on Foundations of Computer Science: October 23–25, 1995, Milwaukee, Wisconsin*, pages 194–203, 1995.

[17] N. Young. The k-server dual and loose competitiveness for paging. *Algorithmica*, 11(6):525–541, 1994.

[18] N. E. Young. On-line file caching. In *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-98)*, pages 82–86, New York, 1998.

[19] Neal E. Young. Bounding the diffuse adversary. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 420–425, 1998.

[20] Neal E. Young. On-line paging against adversarially biased random inputs. 2002.

A Proof of Lemma 2

Proof. For the rest of this proof, we set $P_i = \mathbf{P}[\Gamma(|\sigma|+1) = p \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$ if, provided σ is the prefix, p is the i-th most recently accessed page. Set $|\mathcal{O}(\sigma)| = w$. We next prove that. for $i = 1, \ldots, w$, we have $P_{k+i} \leq 2P_{k+w+i}$. Since $\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] = \sum_{i=1}^{w} P_{k+i}$ and $\mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] = \sum_{i=1}^{w} P_{k+w+i}$, the thesis follow. By the definition of concavity we immediately have that $P_{k+w+i} - P_{k+2w+i} \geq P_{k+i} - P_{k+w+i}$. Observe that the inequality above is always well defined, since $|\mathcal{P}| \geq 4k$. On the other hand, if $P_{k+i} > 2P_{k+w+i}$, the inequality above implies: $P_{k+w+i} - P_{k+2w+i} \geq P_{k+i} - P_{k+w+i} > P_{k+w+i}$. This in turn implies $P_{k+2w+i} < 0$, a contradiction. The lemma is thus proved. □

B Proof of Lemma 6

Proof. For the rest of this proof, we set $P_i = \mathbf{P}[\Gamma(|\sigma|+1) = p \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$ if, provided σ is the prefix, p is the i-th most recently accessed page. Set $|\mathcal{O}(\sigma)| = w$. As an obvious consequence of the definition of convexity, we have, for every $m = 1, \ldots, |\mathcal{P}| - w$: $P_m - P_{m+w} \leq P_{m-w} - P_m$. Now set:

$$\begin{aligned} W_{-1} &= \sum_{i=1}^{w} P_{k+w+i} \leq \mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{N}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] \\ W_0 &= \sum_{i=1}^{w} P_{k+i} = \mathbf{P}[\Gamma(|\sigma|+1) \in \mathcal{O}(\sigma) \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma] \end{aligned}$$

and, for $l = 1, \ldots, t$:

$$W_l = \sum_{i=1}^{w} P_{k-wl+i}.$$

By the convexity of $\mathbf{P}[\Gamma(i) = p \,|\, \overline{\Gamma}(|\sigma|+1) = \sigma]$ we have, for $l = 1, \ldots, t$: $W_l - W_{l-1} \geq W_{l-1} - W_{l-2}$, hence $W_l \geq (l+1)(W_0 - W_{-1})$. This implies:

$$\sum_{l=1}^{t} W_l \geq \sum_{i=1}^{t} (l+1)(W_0 - W_{-1}) \geq \frac{t^2}{2}(W_0 - W_{-1}).$$

If $W_0 > 2W_{-1}$, the above implies $\sum_{l=1}^{t} W_l > \frac{t^2}{4} W_0$. This proves the claim. □