Online paging

Resource augmentation

Idea: To compare the number of misses for an algorithm with the minimal no. of misses for a smaller cache.

For a seq. $\sigma = \sigma_1 \sigma_2 \cdots$ and an integer h, let.

$c_k(A, \sigma)$: No. of misses by A on σ when the cache size is k.

$$\frac{c_k(A, \sigma)}{c_h(A_{opt}, \sigma)} = ? \qquad h < k.$$

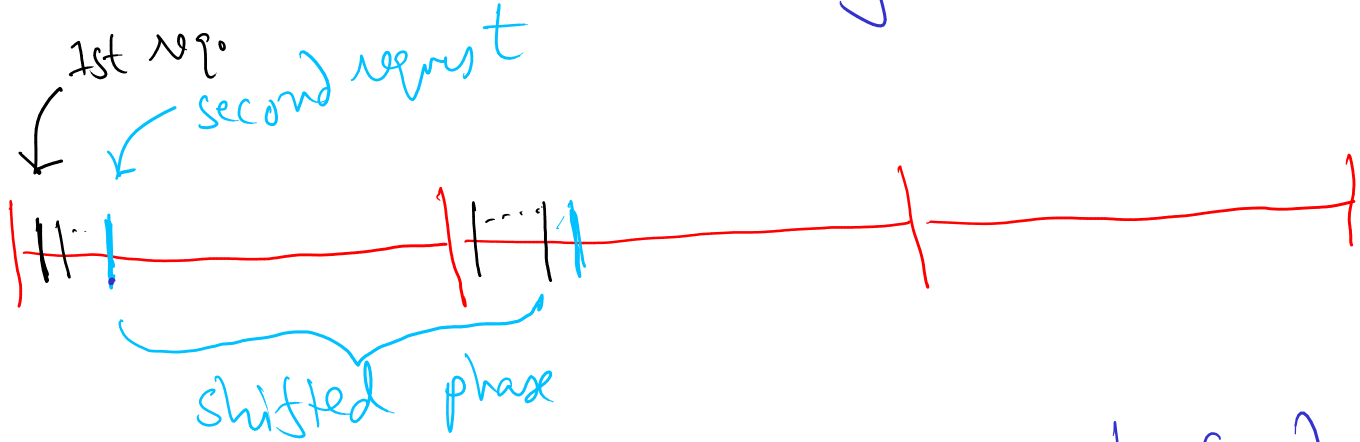
Theorem: If A is a conservative algorithm then there exists a constant b s.t. for every $h \leq k$, and for every σ

$$c_k(A, \sigma) \leq \frac{k}{k-h+1} \cdot c_h(A_{opt}, \sigma) + b.$$

Setting $h = \frac{k}{2}$ yields a factor of ~ 2.

<u>Proof</u>: (1) For each phase A incurs at most k misses (from the definition of a conservative algorithm)

(2)



Just before the beginning of a shifted phase, A_{opt} has the first request of the phase, say p, in its cache. In the shifted phase, there are k requests different from p. But, before the beginning of the shifted page, A_{opt} has $h-1$ pages different from p in its cache. So, it must incur at least $k-h+1$ misses.

<u>Randomized paging algorithms</u>

Algorithm has the freedom of making decisions randomly.

Oblivious adversary model (OA)

An algorithm A has a competitive ratio of α in OA model if there exists a constant b s.t. $\forall \sigma$,

$$\mathbb{E}\left[C(A,\sigma)\right] \leq \alpha \cdot C(A_{OPT}, \sigma) + b.$$

Other adversary models:

As good as deterministic!!

- ~~Fully adaptive adversary~~: Has access to the random bits that the algorithm uses.

- Adaptive adversary: while generating a request, the adversary knows answers to all previous requests.

Marking algorithms: -

- In the beginning of a phase, all pages in the cache are unmarked.
- When a request σ arrives
  - if $\sigma \notin$ Cache, evict an unmarked

page.

— Mark σ

example: LRU

Exercise (1) Verify that LRU is a marking algorithm and FIFO is not a marking algorithm.

(2) Any deterministic marking algorithm is k-competitive.

The algorithm MARK (a randomized marking algorithm)

- Evicts an unmarked page uniformly at random.

Some observations about marking algorithms.

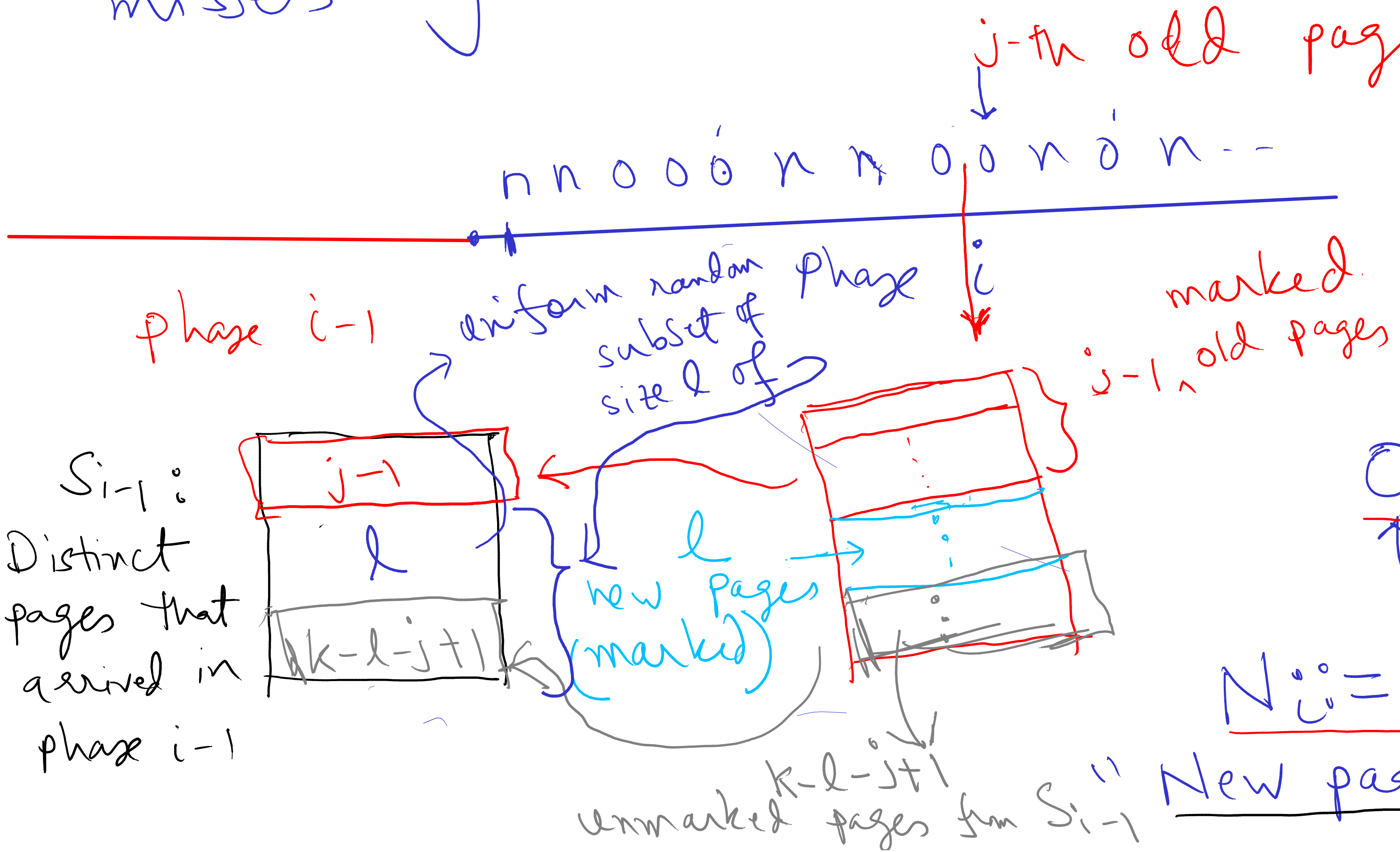
(1) Phases are determined by σ. In particular, phases are independent of the actions of the algorithm.

(2) In each phase, each page causes fault at most once.

(3) After each phase, the cache contains exactly the k distinct pages that arrived in the phase.

<u>Theorem</u>: MARK is $2H_k$-competitive.

<u>Proof</u>: <u>Step 1</u>: Upper-bounding the number of misses by MARK in a phase $i > 1$



S_i: Distinct pages that arrived in phase i.

$\underline{O_i := S_i \cap S_{i-1}}$ "<u>old pages</u>"

$\underline{N_i := S_i \setminus O_i}$ "<u>New pages</u>"

$m_i := |N_i|$ $\qquad S_{i-1} = O_i \cup N_i$

Number of misses on new pages $= m_i$

Number of misses on old pages:

Consider the j-th distinct old page requested.
Before this ℓ new pages (say) were requested

$\ell \le m_i$. $\quad j \le k - m_i$.

$$\Pr\left[\text{MARK incurs a miss on the } j\text{-th distinct old page requested}\right] = \frac{\ell}{k-j+1} \le \frac{m_i}{k-j+1}.$$

Expected number of misses incurred on old pages in phase i:

For $j = 1, \ldots, k - m_i$

$$I_j = \begin{cases} 1 & \text{if the } j\text{-th distinct old page request is a miss,} \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$ Number of misses on old pages in phase i

$$= \sum_{j=1}^{k-m_i} I_j .$$

$\therefore$ Expected no. of such misses $= \sum_{j=1}^{k-m_i} \mathbb{E}[I_j]$

$$= \sum_{j=1}^{k-m_i} \Pr[I_j = 1] \leq \sum_{j=1}^{k-m_i} \frac{m_i}{k-j+1}$$

$$= m_i \left[\frac{1}{m_i+1} + \cdots + \frac{1}{k} \right] = m_i \left(H_k - H_{m_i} \right).$$

$\therefore$ Expected number of misses in phase i

$$\leq m_i \left(1 + H_k - H_{m_i} \right) \leq m_i \cdot H_k .$$

$\therefore$ Expected total number of misses $\leq \left(H_k \cdot \sum_{i=1}^{p} m_i + k \right)$

[p : Total number of phases].

<u>Step 2</u> : Lower-bounding number of misses by an optimal answer sequence.

Phase $i-1$ Phase i

Total number of distinct requests in phases $i-1$ & i $= 2k - |O_i| = k + |S_i| - |O_i|$

$= k + |N_i| = k + m_i$

$\therefore$ An optimal answer sequence incurs at least m_i misses.

No. of misses by an opt. seq. in phase i

$=: F_i$

$$F_{i-1} + F_i \geq m_i \qquad (i \geq 2).$$

$$\Rightarrow \frac{1}{2} \cdot F_{i-1} + \frac{1}{2} F_i \geq \frac{1}{2} m_i$$

$$\sum_{i=2}^{p} \left(\frac{1}{2} \cdot F_{i-1} + \frac{1}{2} F_i \right) \geq \frac{1}{2} \sum_{i=2}^{p} m_i$$

$$\Rightarrow \frac{1}{2} F_1 + \sum_{i=2}^{p-1} F_i + \frac{1}{2} F_p \geq \frac{1}{2} \cdot \sum_{i=2}^{p} m_i$$

$\Rightarrow$ Total no. of misses by an opt. seq.

$$\geq \sum_{i=2}^{p-1} F_i \geq \frac{1}{2}\cdot\sum_{i=2}^{p} m_i - k$$

$$\boxed{\frac{\mathbb{E}[\#\text{ misses by MARK}]}{\#\text{ misses by an opt. seq}} \leq \frac{H_k\cdot\sum_{i=2}^{p} m_i + k}{\frac{1}{2}\cdot\sum_{i=2}^{p} m_i - k}}$$

$$\Downarrow$$

$$\mathbb{E}[\#\text{ misses by MARK}] \leq H_k\cdot\sum_{i=2}^{p} m_i + k$$

$$= 2H_k\left(\frac{1}{2}\cdot\sum_{i=2}^{p} m_i - k\right) + \underbrace{2k\cdot H_k + k}_{b}$$

$$\leq 2H_k\cdot(\#\text{ misses by an opt. seq}) + b$$

$\Rightarrow$ Competitive ratio of MARK is at most $2H_k$.