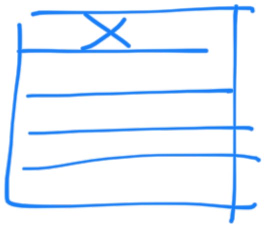
Online Caching.

Input:

$\sigma_1 \sigma_2 \sigma_3 \ldots$

$\sigma_1 \sigma_2 \ldots \sigma_n$

p



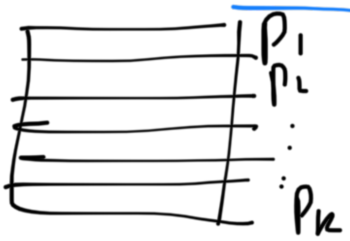
Measure how good an online alg. is?

#evictions → Alg(I)

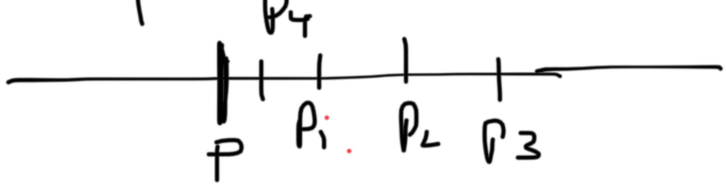
$I = \sigma_1 \ldots \sigma_t \, \sigma_n$

$$\sup_I \frac{Alg(I)}{Opt(I)}$$

p



competitive ratio



① $n = K+1$ pages

K, n ↑ #pages

A

adversary

$n \gg K$.

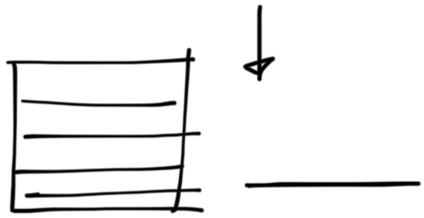
$I: \sigma_1 \sigma_2 \ldots \sigma_T$

K

σ_t : request for a page not in the cache

$Alg(I) = T$

$Opt(I) = T/(K-1)$



K K-1

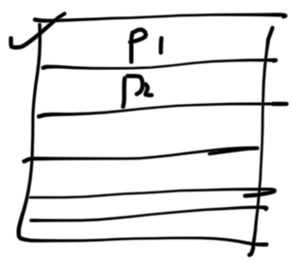
Competitive ratio ≥ K

Q: In general, can we achieve competitive ratio ≤ K? Yes, K

Q: Does randomization help? ← ln K

Alg: LRU   1-bit LRU   K

marked/unmarked



Initially all pages are unmarked.

At time t, P_t is requested

(1) P_t is in the cache: mark P_t

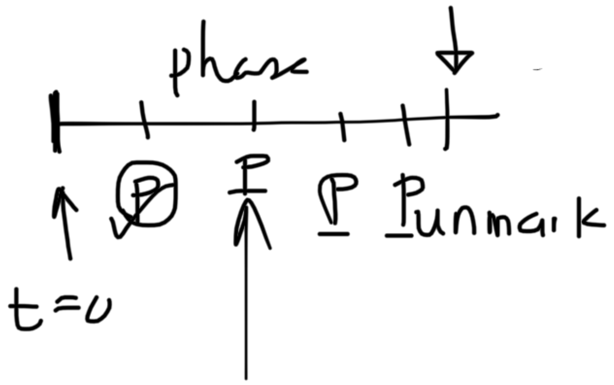
(2) P_t is not in the cache:

② If

- evict ~~an~~ randomly chosen unmarked page
- bring in p_t and mark it

What if all pages are marked?

✓ UnMark all the pages



phase starts when all pages are unmarked

During a phase ≤ k cache misses happen.

\# cache misses ≤ k · # phases ✓

During a phase: k+1 pages are accessed

Cost of any alg ≥ # phases ✓

Alg

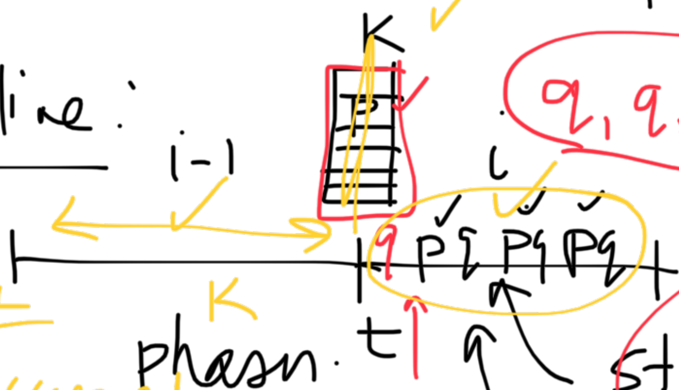
$I = \sigma_1 \dots \sigma_n$

$$\frac{E_I[Alg]}{opt(I)}$$

Expected C.R. on I

Yao's lemma: Exp. C.R. $\geq \ln K$

Outline:



$q_1, q_2 \cdots q_{n_i}$ $\quad p\, p'$ stale

S_i stale request

p new page req. n_i

total # pages accessed

$= K + \frac{n_i}{n_i} + \left[\frac{n_i}{K} + \frac{n_i}{K-1} + \cdots \right]$

$\leq n_i + n_i \ln K \approx n_i \ln K$

Expected # cache misses

$$\sum_i (1+n_i) \ln K$$

phase $i-1, i$: $\geq n_i$ cache misses

$i+1, i+2$: $\geq n_{i+2}$

$\geq n_1 + n_3 + n_5 + \cdots$

$\geq n_2 + n_4 + n_6 + \cdots$

$\frac{1}{2}[n_1 + n_2 + n_3 + \dots]$

Exp. C.R. $\leq O(\ln k)$.

---x---

Extensions :-

(1) Weighted paging : (p) (w_p) ✓

Min. total weighted cache miss.

[BBN '07] : $O(\ln k)$

(2) Generalized paging :

p : s_p, w_p ← cost of evicting p

~~K~~ B ← capacity of cache

total s_p of pages in cache $\leq B$.

[BBN '08] → $O(\log^2 k)$

[ACER '18] → $O(\log k)$

③ Elastic Caching : K, B

K_t

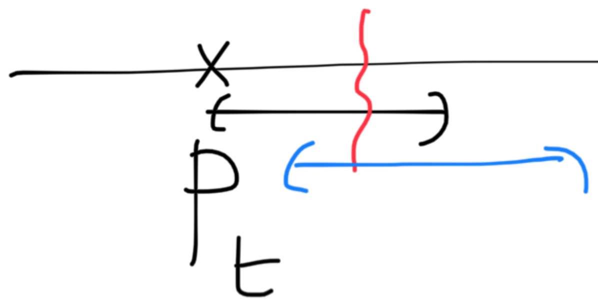
\# cache miss + ~~$\sum_t c(K_t)$~~

$c(s) = 0$ if $s \le K$

∞ if $s > K$

ln n - comp [GPKK'~~20~~ 19]

④ Cache with delays :



[GPK'20]