OSTEP Chapter 22

ECE 3600, Fall 2022

Table of Contents

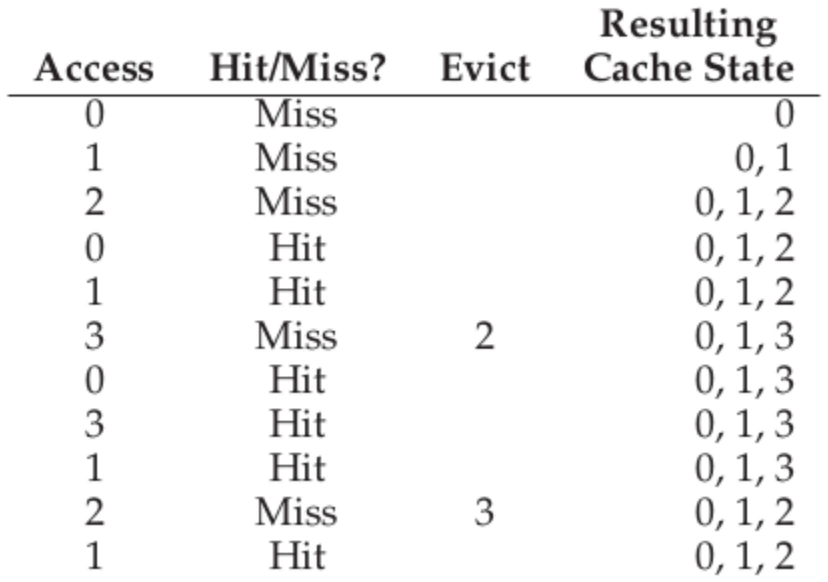
1. Cache Management

Replacement policy to minimize cache misses (maximize cache hits):

when cache is full, replace the page that will be accessed *furthest in the future*.

Requires unrealistic knowledge of the future, but useful for comparisons.

Example, cache size 3, page access: 0, 1, 2, 0, 1, 3, 0, 3, 1, 2, 1



Access	Hit/Miss?	Evict	Resulting Cache State
0	Miss		0
1	Miss		0, 1
2	Miss		0, 1, 2
0	Hit		0, 1, 2
1	Hit		0, 1, 2
3	Miss	2	0, 1, 3
0	Hit		0, 1, 3
3	Hit		0, 1, 3
1	Hit		0, 1, 3
2	Miss	3	0, 1, 2
1	Hit		0, 1, 2

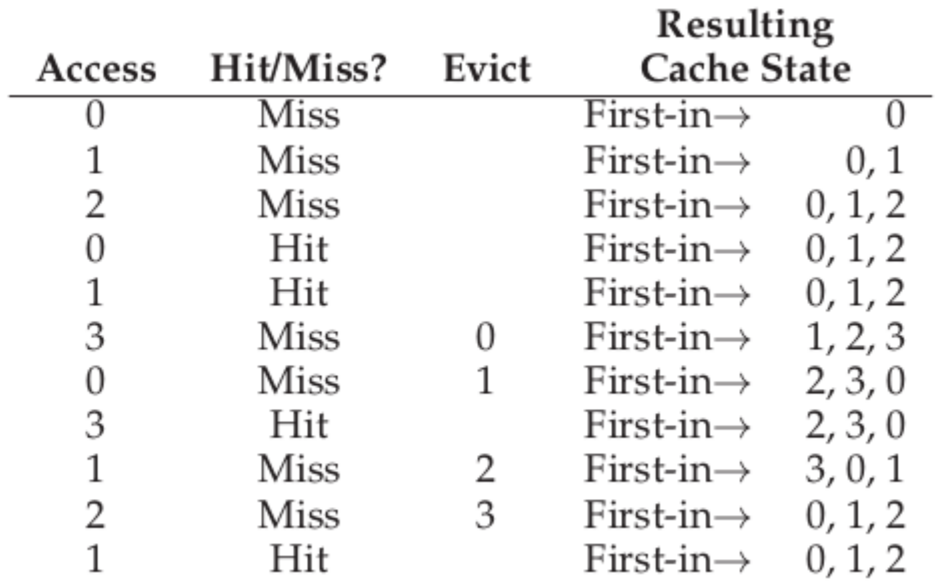
Figure 22.1: **Tracing The Optimal Policy**

6 hits, 5 misses, hit rate = 6/(6+5) = 54.5%

Excluding compulsory misses (first access): 6 hits, 1 miss, hit rate = 6/(6+1) = 85.7%

2. FIFO Policy

first-in, first-out



Access	Hit/Miss?	Evict	Resulting Cache State	
0	Miss		First-in→	0
1	Miss		First-in→	0, 1
2	Miss		First-in→	0, 1, 2
0	Hit		First-in→	0, 1, 2
1	Hit		First-in→	0, 1, 2
3	Miss	0	First-in→	1, 2, 3
0	Miss	1	First-in→	2, 3, 0
3	Hit		First-in→	2, 3, 0
1	Miss	2	First-in→	3, 0, 1
2	Miss	3	First-in→	0, 1, 2
1	Hit		First-in→	0, 1, 2

Figure 22.2: **Tracing The FIFO Policy**

4 hits, 7 misses, hit rate = 4/(4+7) = 36.4%

Excluding compulsory misses: 4 hits, 3 miss, hit rate = 4/(4+3) = 57.1%

3. Random Policy

Access	Hit/Miss?	Evict	Resulting Cache State
0	Miss		0
1	Miss		0, 1
2	Miss		0, 1, 2
0	Hit		0, 1, 2
1	Hit		0, 1, 2
3	Miss	0	1, 2, 3
0	Miss	1	2, 3, 0
3	Hit		2, 3, 0
1	Miss	3	2, 0, 1
2	Hit		2, 0, 1
1	Hit		2, 0, 1

Figure 22.3: **Tracing The Random Policy**

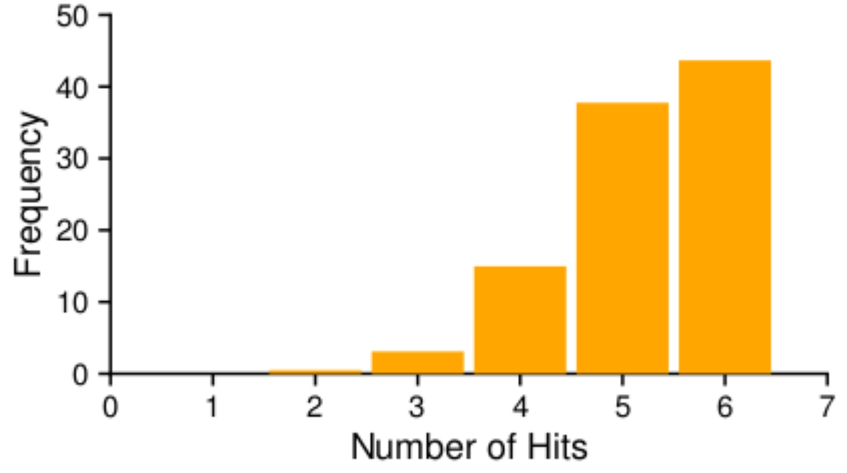
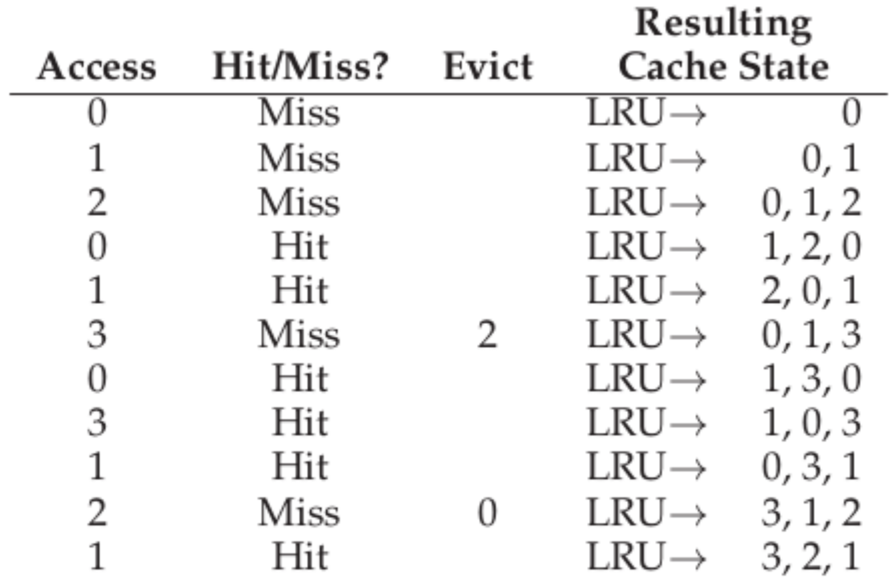


Figure 22.4: **Random Performance Over 10,000 Trials**

4. LRU Policy

consider frequency, recency --> Least-Frequently-Used (LFU), Least-Recently-Used (LRU)



Access	Hit/Miss?	Evict	Resulting Cache State
0	Miss		LRU→ 0
1	Miss		LRU→ 0, 1
2	Miss		LRU→ 0, 1, 2
0	Hit		LRU→ 1, 2, 0
1	Hit		LRU→ 2, 0, 1
3	Miss	2	LRU→ 0, 1, 3
0	Hit		LRU→ 1, 3, 0
3	Hit		LRU→ 1, 0, 3
1	Hit		LRU→ 0, 3, 1
2	Miss	0	LRU→ 3, 1, 2
1	Hit		LRU→ 3, 2, 1

Figure 22.5: **Tracing The LRU Policy**

6 hits, 5 misses, hit rate = 6/(6+5) = 54.5%

Excluding compulsory misses: 6 hits, 1 miss, hit rate = 6/(6+1) = 85.7%

Same hit rate as optimal for this example.

5. No-Locality Workload

each reference is to a random page

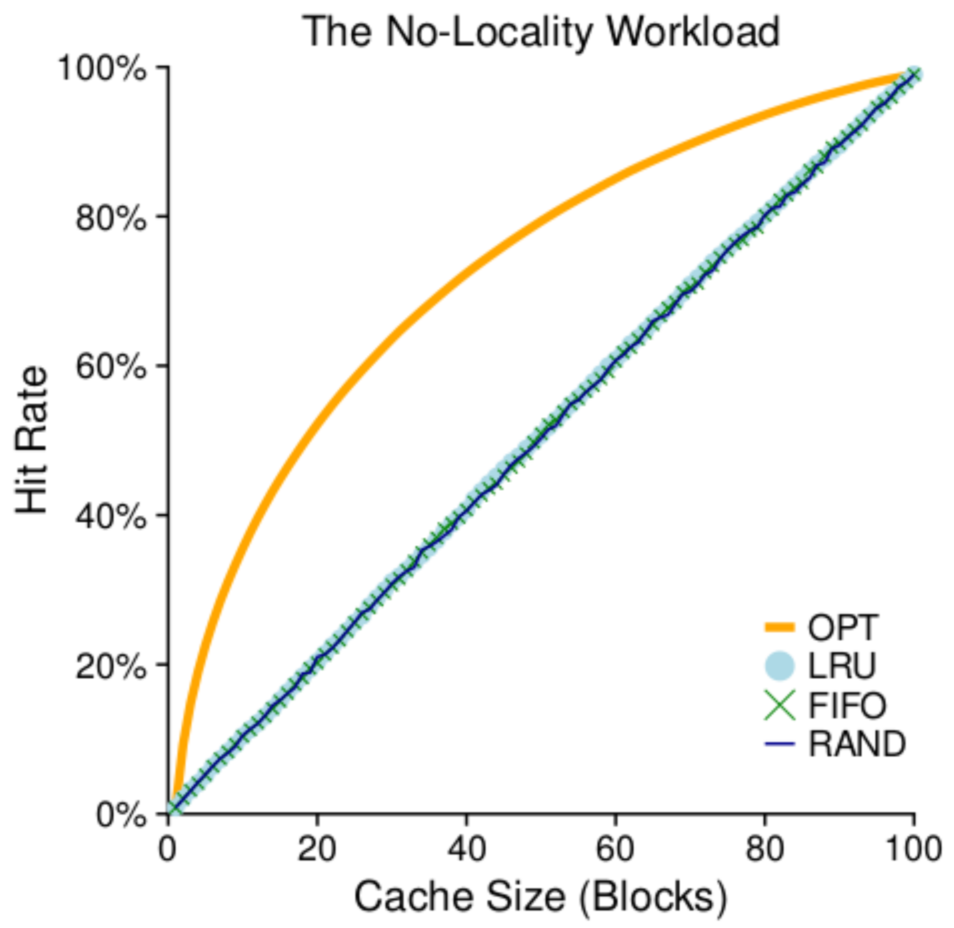


Figure 22.6: **The No-Locality Workload**

6. 80-20 Workload

80% of the references are made to 20% of the pages

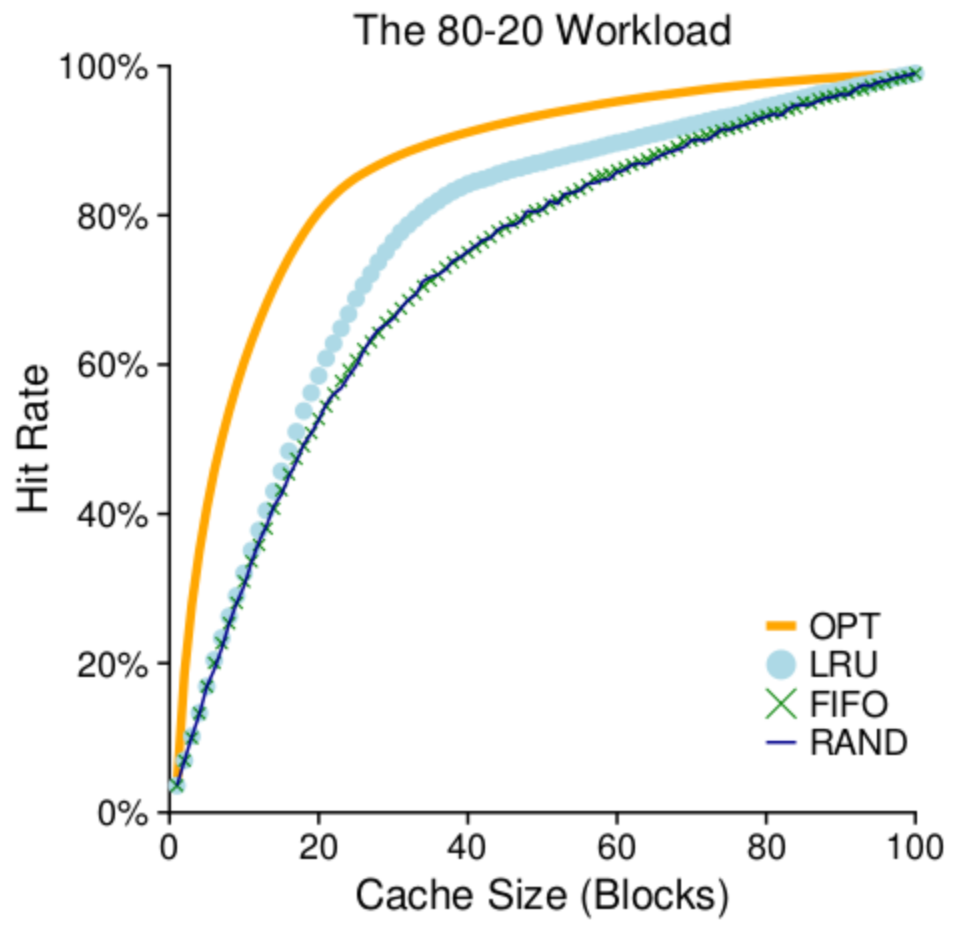


Figure 22.7: **The 80-20 Workload**

7. Looping Sequential Workload

access pages 0, 1, ..., 49, 0, 1, ...

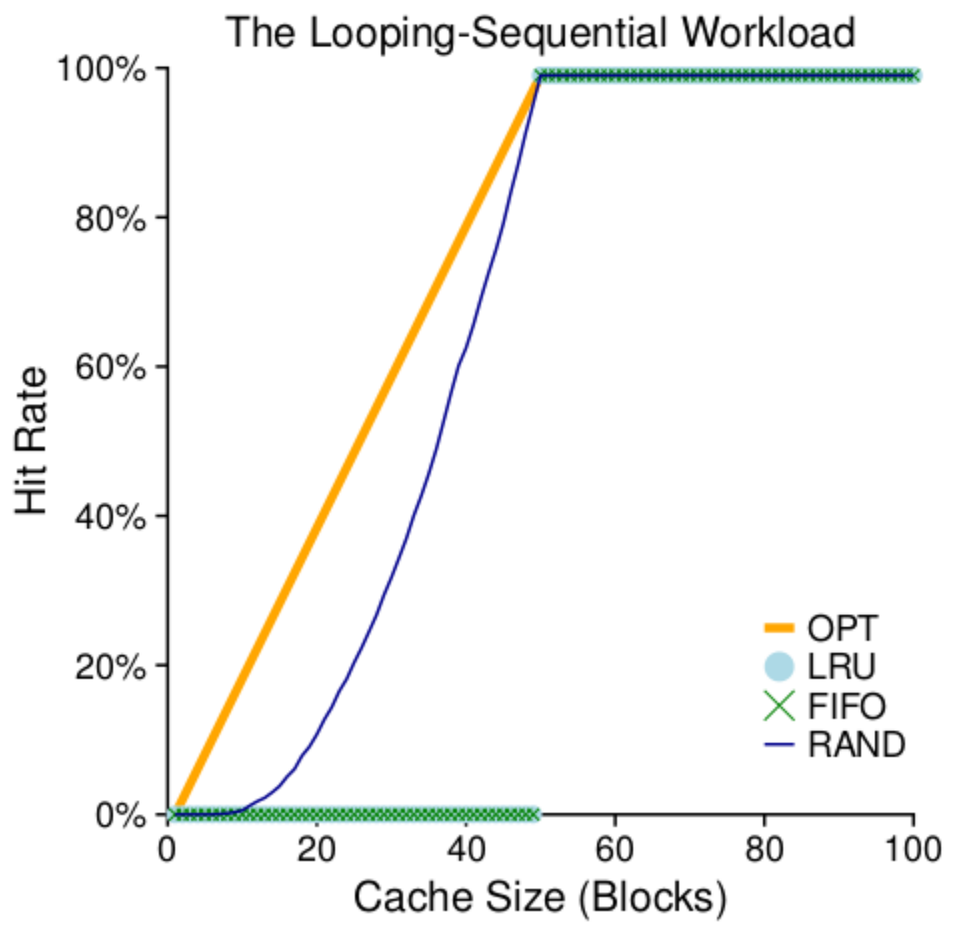


Figure 22.8: **The Looping Workload**

8. Approximating LRU

1-bit "reference bit" and clock algorithm: scan pages, if reference bit is 1, set to 0; else evict.

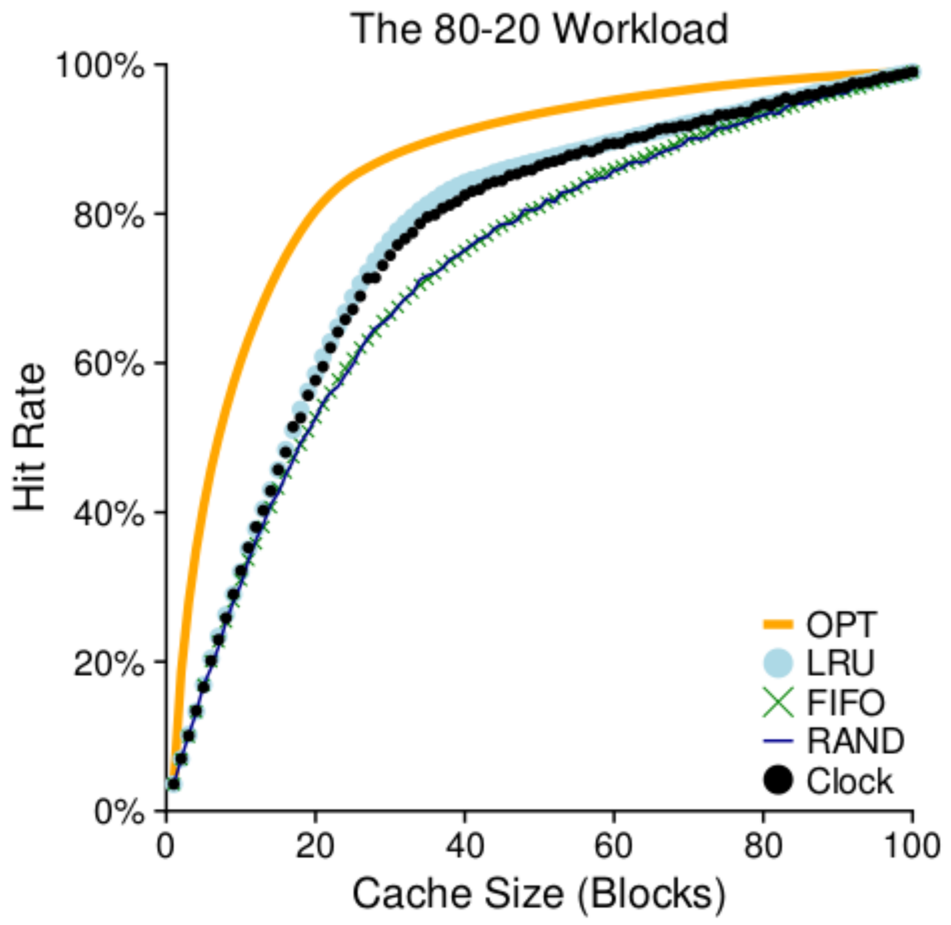


Figure 22.9: **The 80-20 Workload With Clock**

Other considerations: dirty vs. clean pages, demand paging vs. prefetching

9. Exercises

See the book for exercises using paging-policy.py

```
$ python ./paging-policy.py -m 6 -c -s 10
```

```
Access: 3  MISS FirstIn ->          [3] <- Lastin  Replaced:- [Hits:0 Misses:1]
Access: 2  MISS FirstIn ->       [3, 2] <- Lastin  Replaced:- [Hits:0 Misses:2]
Access: 3  HIT  FirstIn ->       [3, 2] <- Lastin  Replaced:- [Hits:1 Misses:2]
Access: 1  MISS FirstIn ->    [3, 2, 1] <- Lastin  Replaced:- [Hits:1 Misses:3]
Access: 4  MISS FirstIn ->    [2, 1, 4] <- Lastin  Replaced:3 [Hits:1 Misses:4]
Access: 4  HIT  FirstIn ->    [2, 1, 4] <- Lastin  Replaced:- [Hits:2 Misses:4]
Access: 3  MISS FirstIn ->    [1, 4, 3] <- Lastin  Replaced:2 [Hits:2 Misses:5]
Access: 0  MISS FirstIn ->    [4, 3, 0] <- Lastin  Replaced:1 [Hits:2 Misses:6]
Access: 3  HIT  FirstIn ->    [4, 3, 0] <- Lastin  Replaced:- [Hits:3 Misses:6]
Access: 1  MISS FirstIn ->    [3, 0, 1] <- Lastin  Replaced:4 [Hits:3 Misses:7]

FINALSTATS hits 3   misses 7   hitrate 30.00
```

Compare with LRU.