

# MS Thesis Defense: Junyao Yang

Friday, October 8 | 3 pm | Rekhi 101

## Efficient Modeling of Random Sampling-Based LRU Cache

The Miss Ratio Curve (MRC) is an important metric and effective tool for caching system performance prediction and optimization. Since the Least Recently Used (LRU) replacement policy is the de facto policy for many existing caching systems, most previous studies on efficient MRC construction are predominantly focused on the LRU replacement policy. Recently, the random sampling-based replacement mechanism, as opposed to replacement relying on the rigid LRU data structure, gains more popularity due to its lightweight and flexibility. To approximate LRU, at replacement times, the system randomly selects  $K$  objects and replaces the least recently used object among the sample.

Redis implements this approximated LRU policy. We observe that there can exist a significant miss ratio gap between exact LRU and random sampling-based LRU under different sampling size  $K$ ; therefore existing LRU MRC construction techniques

cannot be directly applied to random sampling based LRU cache without loss of accuracy.

In this thesis, we present a new probabilistic stack algorithm named KRR which can be used to accurately model random sampling based-LRU cache with arbitrary sampling size  $K$ . We propose two efficient stack update algorithms which reduce the expected running time of KRR from  $O(N \cdot M)$  to  $O(N \cdot \log 2M)$  and  $O(N \cdot \log M)$ , respectively, where  $N$  is the workload length and  $M$  is the number of distinct objects.

Our implementation generates accurate miss ratio curves for both fixed and variable block size cache. Furthermore, we adopt spatial sampling which further reduces the running time of KRR by several orders of magnitude, and thus enables practical, low overhead online application of KRR.



Michigan Tech  
College of Computing

