Khyati Mardia – ASU ID -1215346587

CSE520 Computer Architecture II – Spring 2019
Programming Assignment -2

Different Cache replacement policies :

1. Least Recently Used (LRU) cache replacement policy :

LRU as name suggests discards the least recently used items first and data is inserted at most recently used space. This requires keeping track of what was used when. General implementations of this technique require to keep "age bits" for cache-lines and track the "Least Recently Used" cache-line based on age-bits. In such implementation, every time a cache-line is used, the age of all other cache-lines changes. LRU is sufficient for streaming pattern (no locality of references, infinite reference interval) and works best as no other policy works well. LRU works badly for thrashing patterns (cyclic access pattern (length k , repeat N times)) if k>#blocks in a set. In case of mixed access patterns (some are near immediate re-reference interval, rest are distant), LRU works well for immediate re-reference. It shows low performance with scans (a burst of reference to data whose re-reference interval is in the distant future) where working set does not fit in the cache. Summarizing, LRU works well with workloads with high data locality and works poorly when re-reference occur in the distant future.

2. Re-reference Interval Prediction (RRIP) cache replacement policy :

RRIP predicts the re-reference interval of all missing cache blocks to be an intermediate re-reference interval that is between a near-immediate re-reference interval and a distant re-reference interval. SRRIP is the static RRIP which updates the re-reference prediction to be shorter than the previous prediction upon a re-reference. SRRIP is scan-resistant and thrash resistant. There are two types of SRRIP policies : SRRIP-Hit Priority (SRRIP- HP) and SRRIP-Frequency Priority (SRRIP-FP). SRRIP-HP predicts that any cache block that receives a hit will have a near-immediate re-reference and thus should be retained in the cache for an extended period of time. SRRIP-FP on the other hand predicts that frequently referenced cache blocks will have a near-immediate re-reference and thus they should be retained in the cache for an extended period of time. RRIP works well with mixed access patterns. It uses M-bit re-reference prediction value (RRPV) per cache block. RRPV indicates how soon block will be re-referenced. Whenever there a cache hit, RRPV of block is set to 0. For cache miss, first search for first (2^M-1), if found , replace block and set RRPV to (2^M-2), if not found, increment all RRPVs.

3. Signature-based Hit Predictor (SHiP) cache replacement policy :

Instead of making the same re-reference predications for all cache insertions, distinct signatures are created and are associated with each cache reference. Signature based cache replacement is to predict whether the insertions by a given signature will receive future cache hits. The replacement policy states that if cache insertions by a given signature are re-referenced, then future cache insertions by the same signature will again be re-referenced. Conversely, if cache insertions by a given signature do not receive subsequent hits, then future insertions by the same signature will again not receive any subsequent hits. SHiP has two additional fields - the signature itself and a single bit to track the outcome of the cache insertion. The outcome bit (initially set to zero) is set to one only if the cache line is re-referenced. Signature History Counter Table (SHCT) of saturating counter is incremented when a cache line receives a hit and counter is decremented when a line is evicted from the cache but has not been re-referenced since insertion. Always predicting a near-immediate re-reference interval on cache insertions performs poorly when application references have a distant re-reference interval.

Key differences :

1. LRU has a near immediate re -reference interval ie recently used block is to be used again. LRU predicts that cache blocks are re-referenced in the reverse-order of reference, i.e., LRU predicts that a Most Recently Used (MRU) cache block will be re-referenced much sooner than an LRU cache block.
While LRU provides good performance for workloads with high data locality, LRU limits performance when the prediction of a near- immediate re-reference interval is incorrect. Applications whose re- references only occur in the distant future perform badly under LRU.

2. SRRIP is the re-reference predictions made by RRIP statically on cache hits and misses. SRRIP emulates optimal replacement by correctly predicting a near immediate re-reference interval for the actively used cache blocks and a distant re-reference interval for the scan blocks. RRIP is high performing scan-resistant replacement policy that requires low hardware overhead, retains the existing cache structure, and most importantly integrates easily into existing hardware approximations of LRU. The policy is scan and thrash resistant.

3. Signature-based cache replacement is to predict whether the insertions by a given signature will receive future cache hits. Unlike SRRIP, along with incrementing and decrementing counter on misses and hits, SHiP also predicts a near-immediate re-reference interval on cache insertions. Signature on insertion is recorded and data is re-referenced to the addressed based on cache insertion. SRRIP conservatively predicts that all cache insertions have an intermediate re-reference interval. If the newly-inserted cache line is referenced quickly, SRRIP updates the re-reference prediction to near-immediate; otherwise, SRRIP downgrades the re-reference prediction to distant. In doing so, SRRIP learns the re-reference interval for all inserted cache lines upon re-reference. SHiP on the other hand explicitly and dynamically predicts the re- reference interval based on the SHCT. SHiP makes no changes to the SRRIP victim selection and hit update policies. On a cache miss, SHiP consults the SHCT with the signature to predict the re- reference interval of the incoming cache line.

General Observations :

The results depends on various factors which are as given below.

1. According to paper shared by link, SRRIP improves the application performance by an average of 4-7% over LRU and SHiP improves the application performance by an average of 9.7-9.4% over LRU but we observe that the results are obtained using different set of inputs ie hmmer and mcf (in paper) whereas we are given BFS and MST inputs which may be different from the paper. And hence, similar results are not obtained as mentioned in paper.

2. As the benchmark pattern is different, the results may vary for given inputs

3. The scan pattern can be different for the set of inputs given.

4. The same results are obtained every time as we run the benchmark in gem5 for different cache replacement policies for certain configurations. Performance is not affected because the addresses are maintained the same for gem5 simulator.

5. We are implementing replacement policy as per Hit priority and not the frequency priority. Frequency priority may improve results for certain inputs and configurations.

6. We are setting btp as 100 for SRRIP program which affects the output results.

7. When MST inputs are run on gem5, it is observed that the most tags obtained are in sequential manner. And hence SRRIP, LRU and SHiP outputs for different L2 caches are similar. Whereas BFS input doesn't give sequential tags which changes the output results. And hence, SHiP performs better in case of BFS inputs.

Overall miss rates of L2 Cache for different configurations and policies are shown below

Breadth First Search : I/p file : I/p – BFS_randLocalGraph_J_5_1200000

Overall miss rate in L2 cache for Instructions (in %)

		L2 Cache Size	LRU	SRRIP	ShiP
Breadth First Search	I/p – BFS_randLocalGraph_J_5_120000	256 kB	90.1547%	90.1547%	90.8111%
		1 MB	89.6390%	89.6859%	89.8265%
		4 MB	89.3108%	89.4515%	89.3577%

Overall miss rate in L2 cache for Data (in %)

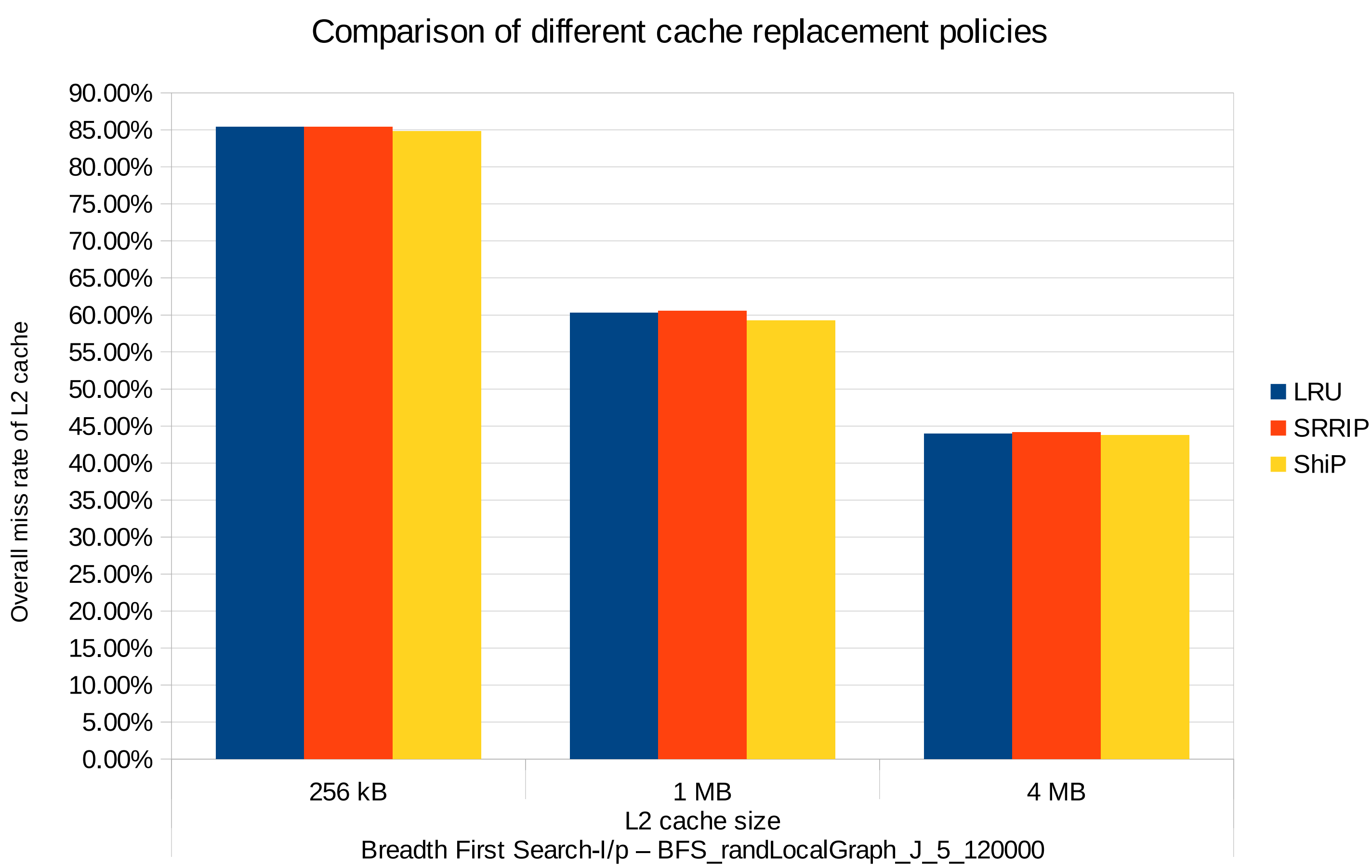
		L2 Cache Size	LRU	SRRIP	ShiP
Breadth First Search	I/p – BFS_randLocalGraph_J_5_120000	256 kB	85.3777%	85.3857%	84.8266%
		1 MB	60.3106%	60.5101%	59.2258%
		4 MB	43.9508%	44.1321%	43.7259%

Total overall miss rate in L2 cache (in %)

		L2 Cache Size	LRU	SRRIP	ShiP
Breadth First Search	I/p – BFS_randLocalGraph_J_5_120000	256 kB	85.3799%	85.3879%	84.8294%
		1 MB	60.3241%	60.5236%	59.2399%
		4 MB	43.9717%	44.1530%	43.7469%

Observations:

The output for SHiP is improved as compared to LRU and SRRIP for all three level of cache for the given input of Breadth First Search. There is no significance difference between LRU and SRRIP for L2 cache size of 256 kB. Where as LRU performs better for 1 MB and 4 MB – L2 cache line size than SRRIP.



Breadth First Search : I/p file : I/p – BFS_rMatGraph_J_5_90000

Overall miss rate in L2 cache for Instructions (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Breadth First Search	I/p – BFS_rMatGraph_J_5_90000	256 kB	90.0563%	90.0563%	90.8537%
		1 MB	89.5403%	89.5872%	89.7749%
		4 MB	89.2589%	89.3058%	89.3527%

Overall miss rate in L2 cache for Data (in %)

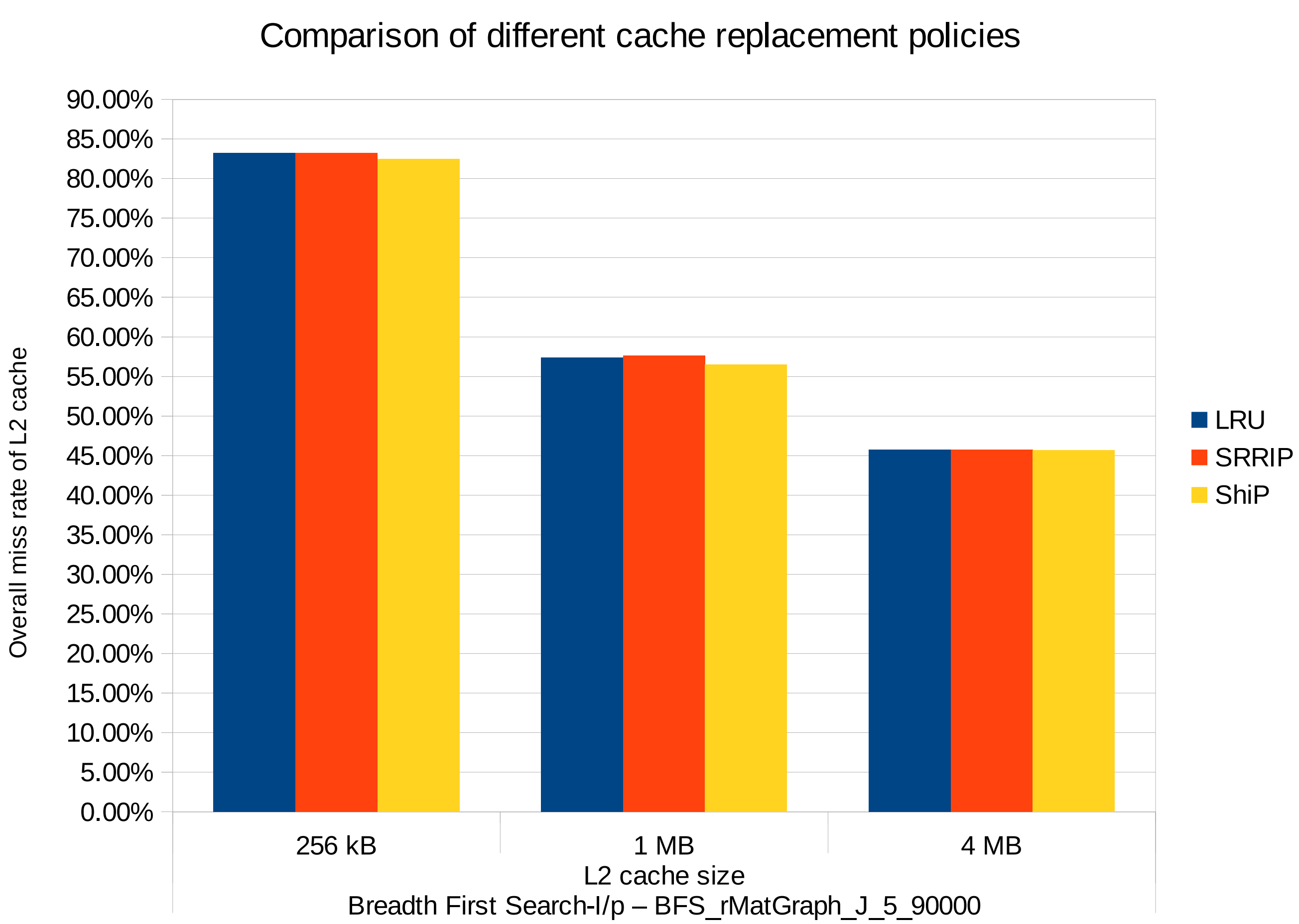
		L2 Cache Size	LRU	SRRIP	SHiP
Breadth First Search	I/p – BFS_rMatGraph_J_5_90000	256 kB	83.2704%	83.2674%	82.5301%
		1 MB	57.4027%	57.6271%	56.4945%
		4 MB	45.7403%	45.7682%	45.7124%

Total overall miss rate in L2 cache (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Breadth First Search	I/p – BFS_rMatGraph_J_5_90000	256 kB	83.2730%	83.2700%	82.5333%
		1 MB	57.4152%	57.6395%	56.5074%
		4 MB	45.7572%	45.7851%	45.7294%

Observations:

The output for SHiP is improved as compared to LRU and SRRIP for 256 kB and 1 MB level of L2 cache for the given input of Breadth First Search. Where as results for SHiP for 4 MB level of L2 cache show similar results as compared to LRU and RRIP. There is no significance difference between LRU and SRRIP for L2 cache size of 256 kB. Where as LRU performs better for 1 MB and 4 MB – L2 cache line size than SRRIP.



Min Spanning Forest : I/p file : I/p – MST_randLocalGraph_WE_5_100000

Overall miss rate in L2 cache for Instructions (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_randLocalGraph_WE_5_100000	256 kB	13.4745%	13.4744%	13.5487%
		1 MB	13.4134%	13.4189%	13.4709%
		4 MB	13.4176%	13.4112%	13.4116%

Overall miss rate in L2 cache for Data (in %)

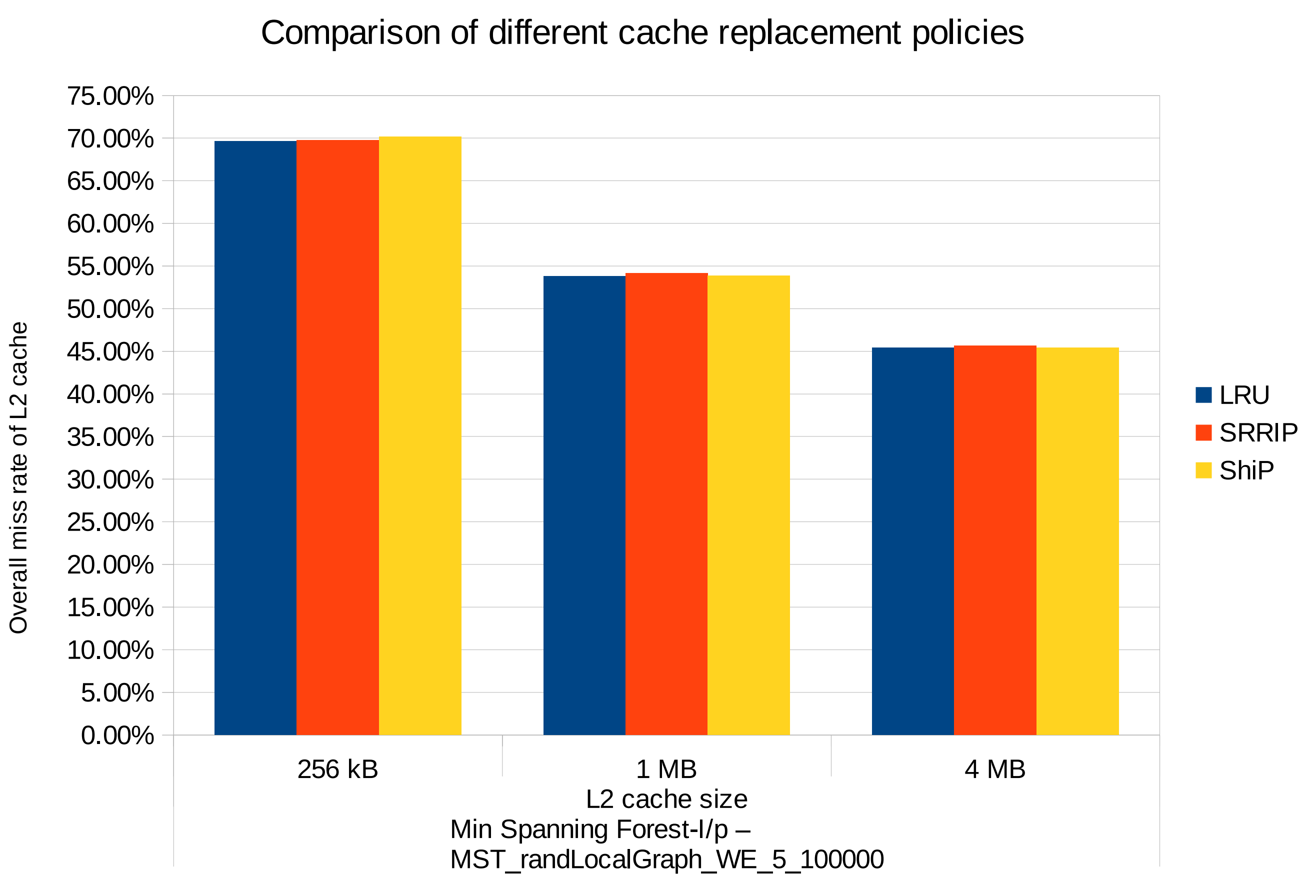
		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_randLocalGraph_WE_5_100000	256 kB	81.1099%	81.2724%	81.7476%
		1 MB	62.6950%	63.0933%	62.7241%
		4 MB	52.9045%	53.1854%	52.8809%

Total overall miss rate in L2 cache (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_randLocalGraph_WE_5_100000	256 kB	69.6272%	69.7667%	70.1747%
		1 MB	53.8235%	54.1654%	53.8487%
		4 MB	45.4215%	45.6625%	45.4012%

Observations:

As tags values are obtained mainly in sequential manner, the best performed policy is LRU for all 3 level of L2 cache. Here, all three cache policies shows similar results. There is no significance difference of SRRIP output as compared to LRU. SHiP performs better for L2 cache size of 1 MB. For all other size ,there is no significance difference.



Min Spanning Forest : I/p file : I/p – MST_rMatGraph_WE_3_35000

Overall miss rate in L2 cache for Instructions (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_rMatGraph_WE_3_35000	256 kB	38.4156%	38.4198%	38.5111%
		1 MB	38.2467%	38.2412%	38.3978%
		4 MB	38.2448%	38.2422%	38.0780%

Overall miss rate in L2 cache for Data (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_rMatGraph_WE_3_35000	256 kB	72.7212%	72.6790%	73.9078%
		1 MB	58.6553%	59.2207%	59.0991%
		4 MB	44.3120%	44.6403%	45.1479%

Total overall miss rate in L2 cache (in %)

		L2 Cache Size	LRU	SRRIP	SHiP
Min Spanning Forest	I/p – MST_rMatGraph_WE_3_35000	256 kB	62.4107%	62.3745%	63.4283%
		1 MB	50.3495%	50.8342%	50.7301%
		4 MB	38.0505%	38.3321%	38.7671%

Observations:

The performance of LRU for all 3 level of cache is slightly better as compared to SRRIP and SHiP. There is no significance difference obtained in SRRIP and SHiP. SHip is better in case of L2 cache level size of 1 MB whereas for the two sizes SRRIP performs slightly better.