Improve Scan Efficiency of SIS

Abel Wu <wuyun.abel@bytedance.com>

Linux Plumbers Conference | Dublin, Ireland Sept. 12-14, 2022

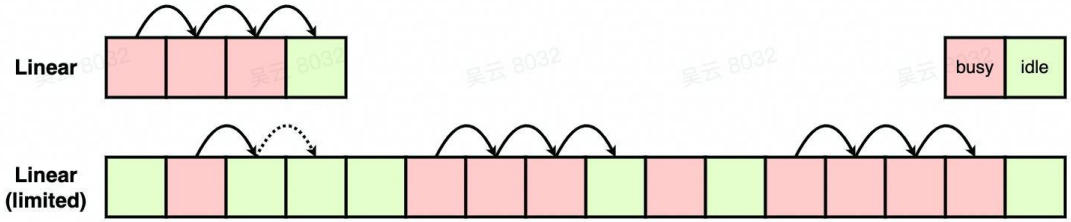
ByteDance 字节跳动

Content

- Background
- Filter
  - False Positive
  - Generation
- Benchmark

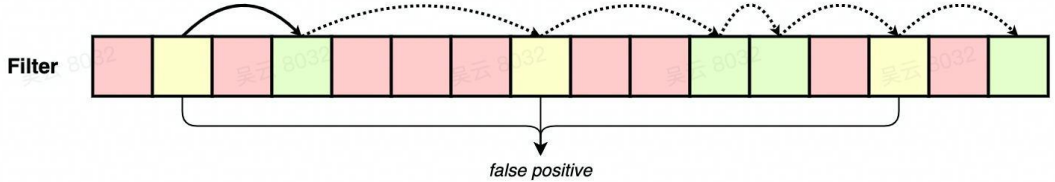
Background

- SIS plays an important role in performance
  - for sufficient use of cpu capacity
  - for better data locality
- SIS now scans linearly which works well when:
  - under light pressure → not hard to find an idle cpu
  - with small LLCs → well bounded worst case



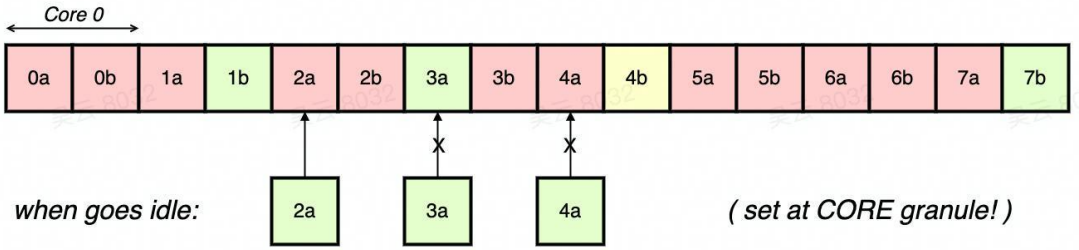
Background

- Trends in the real world
  - CSPs aim at TCO optimization → Co-location
  - Workloads are becoming more complicated and hungry
  - LLCs are getting bigger → How's the scalability?
- Limiting scan depth is not enough
  - SIS_{UTIL,PROP} don't contribute to success rate
  - Would be good to scan more wisely



Filter

- The filter is a cpumask of unoccupied cpus
  - LLC-shared and LLC-specific, resident in shared sched-domain
  - Set cpus at idle path → But when to clear?
  - Set at CORE granule → Lightweight & Balance load

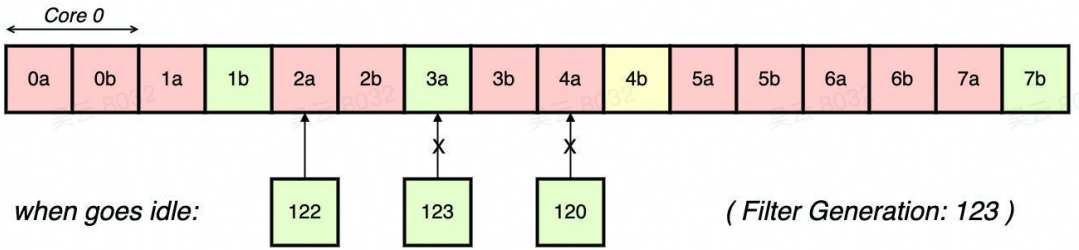


False Positive

- Clear on task enqueue
  - Can be frequent → The filter is LLC-shared
  - Multiple updates before being selected → Not efficient
- Clear periodically on tick/balance/...
  - Can be too late → Wakeup can be way more frequent
  - Not efficient when LLC is not loaded
- Clear when scan starts to fail
  - Adjust when necessary → Lazy & Ondemand
  - Full scan fails → Reset the whole filter
  - Partial scan fails? *TBD*

Generation

- How to know if a cpu is set in filter
  - Straightforward cpumask_test_cpu() → Costly due to LLC-shared
  - Cache locally in runqueue → Need to maintain coherence
- Filter generation
  - Cache generation in runqueue when set cpus to the filter
  - Generation iterates when reset, expiring all caches

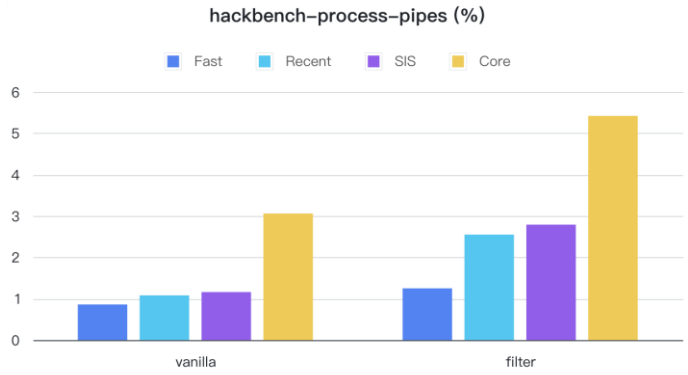


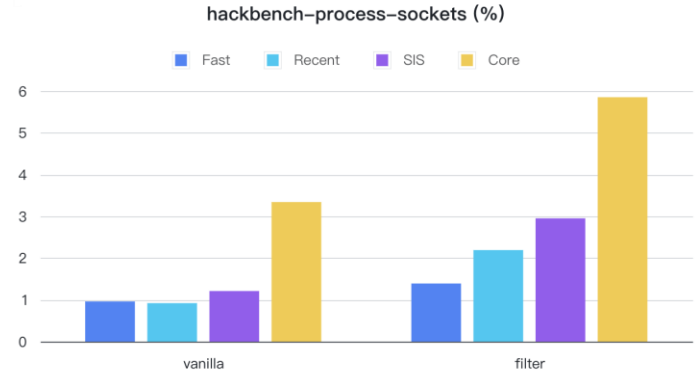
Benchmark

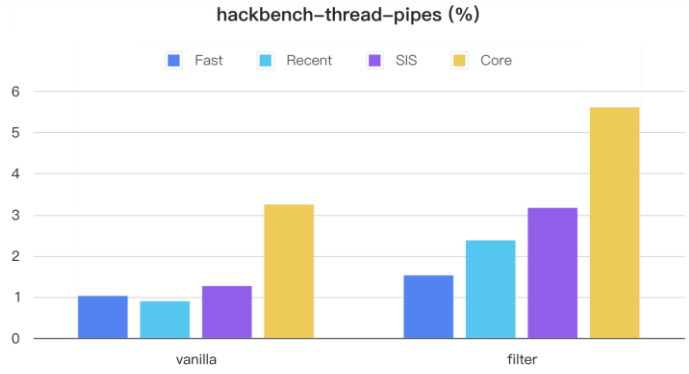
hackbench-process-pipes

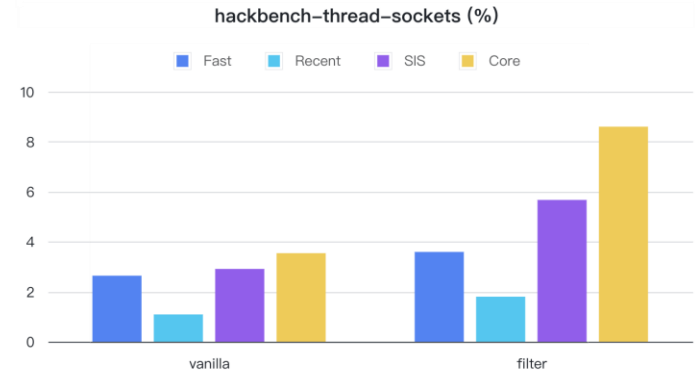
Amean	1	0.2963 (0.00%)	0.2933 (1.01%)		Drew
Amean	4	0.6093 (0.00%)	0.5883 (3.45%)		
Amean	7	0.7837 (0.00%)	0.7570 (3.40%)		
Amean	12	1.2703 (0.00%)	1.0780 (15.14%)		Great Win
Amean	21	2.6260 (0.00%)	1.8903 * 28.01%*		
Amean	30	4.3483 (0.00%)	2.7903 * 35.83%*		
Amean	48	7.9753 (0.00%)	4.8920 * 38.66%*		
Amean	79	9.6540 (0.00%)	8.0127 * 17.00%*		
Amean	110	11.2597 (0.00%)	10.1557 * 9.80%*		Win
Amean	141	13.8077 (0.00%)	12.7387 * 7.74%*		
Amean	172	16.3513 (0.00%)	14.5860 * 10.80%*		
Amean	203	19.0880 (0.00%)	17.1950 * 9.92%*		
Amean	234	21.7660 (0.00%)	19.6763 * 9.60%*		
Amean	265	23.0447 (0.00%)	22.5557 (2.12%)		
Amean	296	25.4660 (0.00%)	24.4273 (4.08%)		

SIS Success Rate









THANKS.

ByteDance字节跳动

tbench4 Throughput

Hmean	1	300.70	(0.00%)	302.52 *	0.61%*
Hmean	2	597.53	(0.00%)	604.76 *	1.21%*
Hmean	4	1188.34	(0.00%)	1204.79 *	1.38%*
Hmean	8	2336.22	(0.00%)	2375.87 *	1.70%*
Hmean	16	4459.17	(0.00%)	4681.25 *	4.98%*
Hmean	32	7606.69	(0.00%)	7607.93 (	0.02%)
Hmean	64	9009.48	(0.00%)	8956.00 *	-0.59%*
Hmean	128	19456.88	(0.00%)	19258.30 *	-1.02%*
Hmean	256	19771.10	(0.00%)	20783.82 *	5.12%*
Hmean	384	20118.74	(0.00%)	20407.40 *	1.43%*

netperf-udp

Hmean	send-64	209.06 (0.00%)	210.32 (0.60%)	
Hmean	send-128	416.70 (0.00%)	415.34 (-0.33%)	
Hmean	send-256	819.65 (0.00%)	808.52 * -1.36%*	
Hmean	send-1024	3163.12 (0.00%)	3132.35 * -0.97%*	
Hmean	send-2048	5958.21 (0.00%)	5926.40 (-0.53%)	
Hmean	send-3312	9168.81 (0.00%)	9194.53 (0.28%)	
Hmean	send-4096	11039.27 (0.00%)	11159.21 * 1.09%*	
Hmean	recv-64	209.06 (0.00%)	210.32 (0.60%)	
Hmean	recv-128	416.70 (0.00%)	415.34 (-0.33%)	
Hmean	recv-256	819.65 (0.00%)	808.52 * -1.36%*	
Hmean	recv-1024	3163.12 (0.00%)	3132.35 * -0.97%*	
Hmean	recv-2048	5958.21 (0.00%)	5926.40 (-0.53%)	
Hmean	recv-3312	9168.81 (0.00%)	9194.53 (0.28%)	
Hmean	recv-4096	11039.27 (0.00%)	11159.10 * 1.09%*	

netperf-tcp

Hmean	64	1192.41 (0.00%)	1219.91 *	2.31%*
Hmean	128	2354.50 (0.00%)	2360.65 (	0.26%)
Hmean	256	4371.10 (0.00%)	4393.92 (	0.52%)
Hmean	1024	13813.84 (0.00%)	13712.10 (	-0.74%)
Hmean	2048	21518.91 (0.00%)	21950.82 *	2.01%*
Hmean	3312	25585.77 (0.00%)	26087.72 *	1.96%*
Hmean	4096	27402.77 (0.00%)	27927.67 *	1.92%*
Hmean	8192	31766.67 (0.00%)	31914.49 (	0.47%)
Hmean	16384	36227.30 (0.00%)	36630.26 (	1.11%)