Observations on mTHP alloc&swpout fallbacks

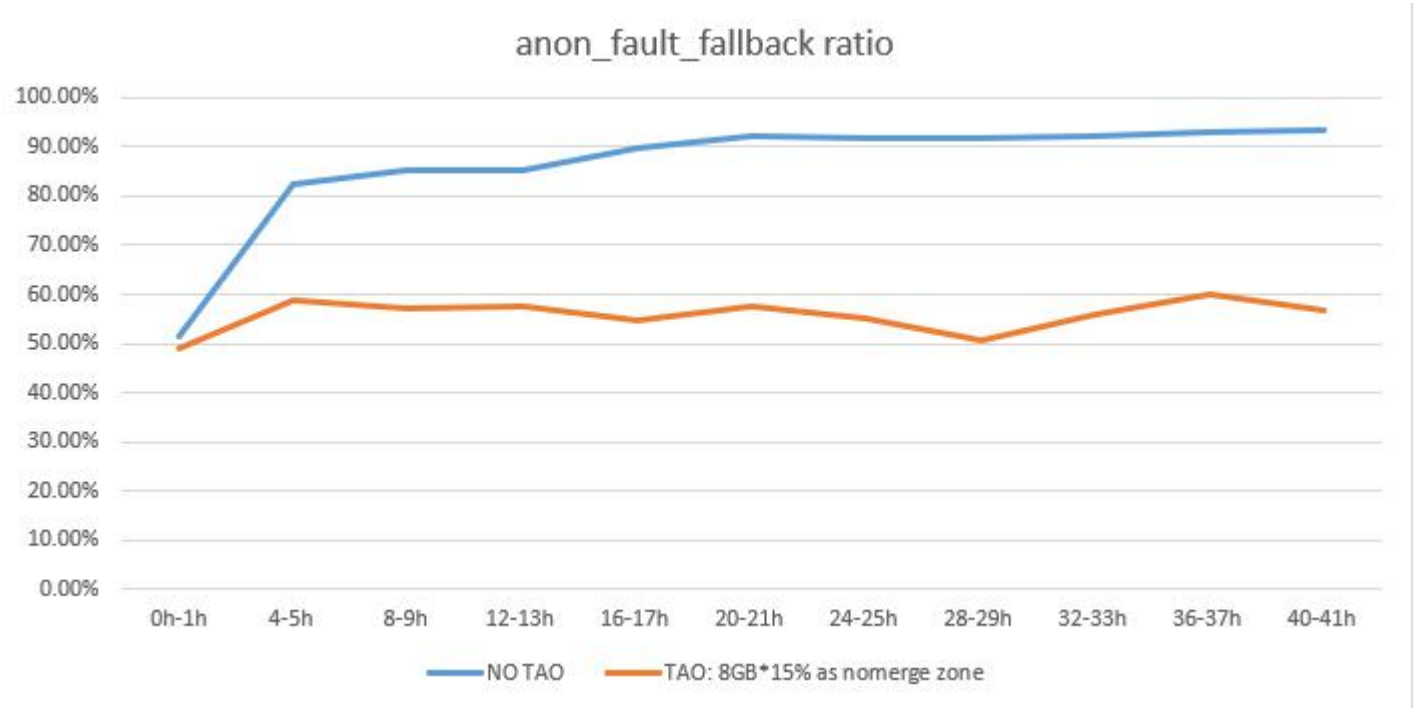
oppo

zram_6g_swap50%(loop am_start)													
	NO TAO					NO TAO					TAO: 8GB*15% as nomerge zone			
	0h	1h	0h-1h	fallback ratio		0h	1h	0h-1h	fallback ratio		0h	1h	0h-1h	fallback ratio
anon_fault_alloc	2418	1259912	1257494			3266	1498085	1494819			3381	1463983	1460602	
anon_fault_fallback	0	1329616	1329616	51.39%		0	1929688	1929688	56.35%		0	1407949	1407949	49.08%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	0	220148	220148			0	199067	199067			0	116271	116271	
anon_swpout_fallback	0	28296	28296	11.39%		0	45863	45863	18.72%		0	12781	12781	9.90%
	4h	5h	4-5h			4h	5h	4-5h	fallback ratio		4h	5h	4-5h	
anon_fault_alloc	3445037	3906438	461401			4115735	4795330	679595			5678769	6981737	1302968	
anon_fault_fallback	6828576	8970439	2141863	82.28%		9888584	12688125	2799541	80.47%		6682142	8531995	1849853	58.67%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	287413	290813	3400			262182	272436	10254			147609	151063	3454	
anon_swpout_fallback	338399	414790	76391	96%		312945	391871	78926	89%		60766	71417	10651	76%
	8h	9h	8-9h			8h	9h	8-9h	fallback ratio		8h	9h	8-9h	
anon_fault_alloc	5206787	5628911	422124			6538124	6959766	421642			10823713	12141712	1317999	
anon_fault_fallback	15205978	17616091	2410113	85.10%		21050135	23961914	2911779	87.35%		13416021	15165662	1749641	57.04%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	296553	297885	1332			284602	286556	1954			153844	154324	480	
anon_swpout_fallback	629240	688699	59459	98%		611888	667230	55342	97%		98809	105721	6912	94%
	12h	13h	12-13h			12h	13h	12-13h	fallback ratio		12h	13h	12-13h	
anon_fault_alloc	6702713	7087377	384664			7845220	8082313	237093			15656440	16808483	1152043	
anon_fault_fallback	24557459	26778824	2221365	85.24%		33000183	35886983	2886800	92.41%		19943152	21518039	1574887	57.75%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	299173	299435	262			288925	289369	444			155510	155926	416	
anon_swpout_fallback	860299	919356	59057	100%		787127	820193	33066	99%		119988	124340	4352	91%
	16h	17h	16-17h			16h	17h	16-17h	fallback ratio		16h	17h	16-17h	
anon_fault_alloc	8013625	8278911	265286			8698790	8882842	184052			20404400	21554684	1150284	
anon_fault_fallback	33753021	36079007	2325986	89.76%		44367046	47186090	2819044	93.87%		26153343	27531773	1378430	54.51%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	299531	299563	32			290154	290951	797			156566	156566	0	
anon_swpout_fallback	1057620	1098219	40599	100%		903932	929585	25653	97%		135236	136700	1464	100%
	20h	21h	20-21h			20h	21h	20-21h	fallback ratio		20h	21h	20-21h	
anon_fault_alloc	9024577	9222766	198189			9363328	9515028	151700			24960604	26108817	1148213	
anon_fault_fallback	43185898	45500041	2314143	92.11%		55706163	58536472	2830309	94.91%		31975283	33529269	1553986	57.51%
anon_fault_fallback_charge	0	0	0			0	0	0			0	0	0	
anon_swpout	299563	299563	0			291687	291695	8			156598	156598	0	
anon_swpout_fallback	1214433	1244294	29861	100%		1000151	1020257	20106	100%		144867	146828	1961	100%

Unlikely success for mTHP allocations

- **Based on our observations, mTHP proves beneficial only within the initial hour. Subsequently, there's a likelihood of allocation failures.**
- **We have the option to utilize either dedicated page blocks or dedicated virtual zones for large folios.**

the following data was gathered on a Pixel 6 with 8GB of memory:
the success rate of mTHP allocation throughout the entire duration of mobile phone usage is stable.
(Similar data was also observed on OPPO's phone, which utilized dedicated page blocks for large folios.)



Folio Size Conflict in Single LRU

At first, we noticed significant regressions in app launch speed and user experiences following the allocation of specific page blocks for large folios on phones.

LRU could be like

LF - LF - LF - SF - SF - SF - SF - SF - SF -SF - SF - SF - SF - SF - SF - SF

or

SF - SF - SF - LF - LF - LF - LF - LF - LF -LF - LF - LF - LF - LF - LF - LF

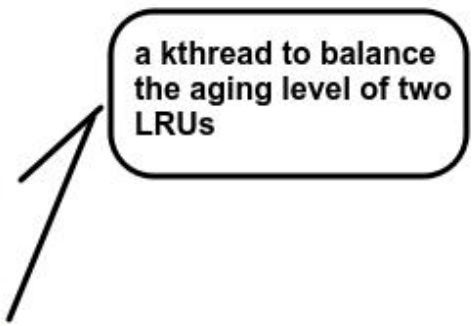
As a result, we opted to implement separate LRUs for large folios in node&memcg, having a kernel thread to balance the aging level of two LRUs

LRU1

SF - SF - SF - SF - SF - SF - SF - SF - SF -SF - SF - SF - SF - SF - SF - SF

LRU2

LF - LF - LF - LF - LF - LF - LF - LF - LF -LF - LF - LF - LF - LF - LF - LF



- The zone-specific LRUs of MGLRU can autonomously address LRU concerns through TAO's virtual zones. MGLRU can balance aging levels across zones.

mTHP Deployed on OPPO's Android Phones