



Contribution ID: 285

Type: not specified

Unifying sources of page hotness information

In the Linux kernel, hot page information can potentially be obtained from multiple sources:

- a. `PROT_NONE` faults (NUMA balancing)
- b. PTE Access bit (LRU scanning)
- c. Hardware provided page hotness info (AMD IBS, CXL HMU)

This information is further used to migrate (or promote) pages from slow to top memory tier for optimal performance.

Currently, the sources a) and b) above work independently of each other while integrating the information provided by sources mentioned in c) is in the development stage [1]. Apart from the above major sources, there are other instances within the kernel where memory access information is detected and acted upon. With such multiple sources of page temperature information being available, it makes sense to collate information from all of them and maintain a single source of page hotness information in the kernel. Sub-systems (like hot page promotion) that act on page hotness information can make use of this information for further action.

This proposal builds on a similar proposal that was discussed in LSFMM [2] early this year. It aims to continue the discussion further and look at the challenges with the patchset [3] proposed for the same.

Discussion points:

- The feasibility, pros and cons of having a centralized access tracking subsystem within the kernel.
- The overheads due to access tracking.
- The optimal hot page access record representation.
- The suitability and availability of different page hotness sources.

[1] <https://lists.gnu.org/archive/html/qemu-devel/2025-06/msg02196.html>

[2] <https://lwn.net/Articles/1016722/>

[3] <https://lore.kernel.org/linux-mm/20250910144653.212066-1-bharata@amd.com/>

Primary author: RAO, Bharata Bhasker (AMD)

Presenter: RAO, Bharata Bhasker (AMD)

Session Classification: Device and Specific Purpose Memory MC

Track Classification: Device and Specific Purpose Memory MC