



Contribution ID: 211

Type: **not specified**

Reducing the number of users of mmap_sem

Wednesday, 14 November 2018 12:00 (15 minutes)

The `mmap_sem` has long been a contention point in the memory management subsystem. In this session some `mmap_sem` related topics will be discussed. Some optimization has been merged by the upstream kernel to solve holding `mmap_sem` for write for excessive period of time in `munmap` path by downgrading write `mmap_sem` to read. And, some optimization are under discussion on the mailing list, i.e. release `mmap_sem` earlier for page cache readahead, speculative page fault. There is still optimization room by figuring out just what `mmap_sem` protects. It covers access to many fields in the `mm_struct` structure. It is also used for the virtual memory area (VMA) red-black tree, the process VMA list, and various fields within the VMA structure itself. Finer grain locks might be better to replace `mmap_sem` to reduce contention, i.e. range lock or per vma lock.

I agree to abide by the anti-harassment policy

Presenter: SHI, Yang (Alibaba Group)

Session Classification: Performance and Scalability MC