Week 10

Paper Critique

**Efficient Virtual Memory for Big Memory Servers (ISCA '13)**

This paper suggests the direct segment and primary region to support direct mapping between virtual address and physical address to avoid TLB lookup which is an serious overhead in big memory workload.

The advantage of suggested paper is that it efficiently address the properties of the big memory workload and solve out the TLB MISS overhead that eventually cause page table traverse by avoid the TLB lookup. However, the limitation that lies in this paper is that this kind of method only applies well in big memory workload, but not the best solution for in general cases. Moreover, in a server running dynamic workload including the big memory workload would not fit well.

The improvement that can be done here is considering the more general cases and changes direct segment size more dynamically but efficiently. Since the fixed size of direct segment does not work well so the paper suggests that the size would change time to time through monitoring, but the workload that benefits well with the short term keep of direct segment will cause intervention.

**Coordinated and Efficient Huge Page Management with Ingens (OSDI '16)**

This paper proposes a novel framework called Ingens for huge page support that relies on a basic primitives to provide transparent huge page support. It handles page fault with asynchronous allocation and monitors spatial utilization of each huge page region and allocates pages based on that.

The contribution of this paper is that it does not incur extra page fault latency and bound memory bloating which was overhead in conventional scheme. Huge page can benefit when the workload makes use of contiguous big memory.

However, in a small randomized access pattern with small update huge page may not fit well which is the reason for not using huge page. Considering more fairness way of allocating a huge page with coordinated memory management would improve Ingens.