Efficient Virtual Memory for Big Memory Servers, ISCA 2013

There are many moves to alleviate TLB missies, which can account for upto 51% of execution time in Big memory workloads.

To remove the TLB miss overhead for big-memory workloads, this paper proposed **Direct Segments.** If a virtual address is In the range of direct segment(BASE~LIMIT), virtual address could be interpret to mapped to physical address just by adding offset value(OFFSET). It means ther is no need to lookup TLB. They add three register to store values BASE, LIMIT, OFFSET.

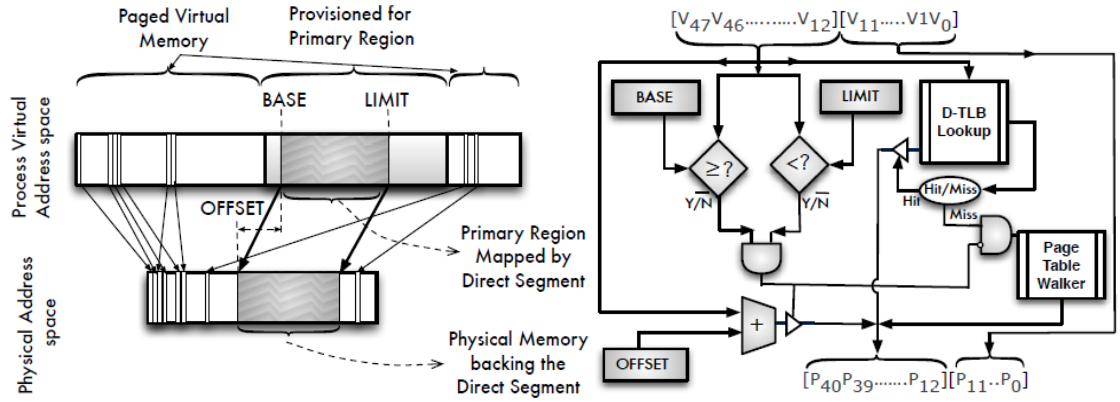
They made a software abstraction Primary Region, which expect the use of direct segment propose mapping part of a process's linear virtual address space with a direct segment, while page mapping the rest of the virtual address space.

Pros

While keeping conventional page-based virtual memory for compatibility in outer-range of primary region. it effectively reduce TLB misses using linear direct mapping between virtual memory and physical memory

Cons

Sparse virtual memory allocations may waste physical memory if mapped with direct segments.



Coordinated and Efficient Huge Page Management with Ingens, OSDI 2016

As address translation cost get high, people try to solve them by using huge pages (expect to alleviate TLB missis). However, Current huge page management suffer several problems.

Think of allocating huge pages. If memory does not have enough contiguous memory, linux compact physical memory and it can induce long tail latency. Ingen solve this by asynchronously allocate huge pages

Also by changing allocating mechanism (from greedy allocation) to spatial utilization based allocation, it gives bounded memory bloating

Pros

Ingens provides principled, coordinated transparent huge page support for the operating system and hypervisor.

Cons / ideas

Paper do not have NUMA evaluation and considerations. Considering them and adding inspect of memory share situations across NUMA would be better.