Advanced Parallel Memory Virtualization

Zhang Yulei <yuleixzhang@tencent.com>

Tencent Cloud

Tencent Cloud

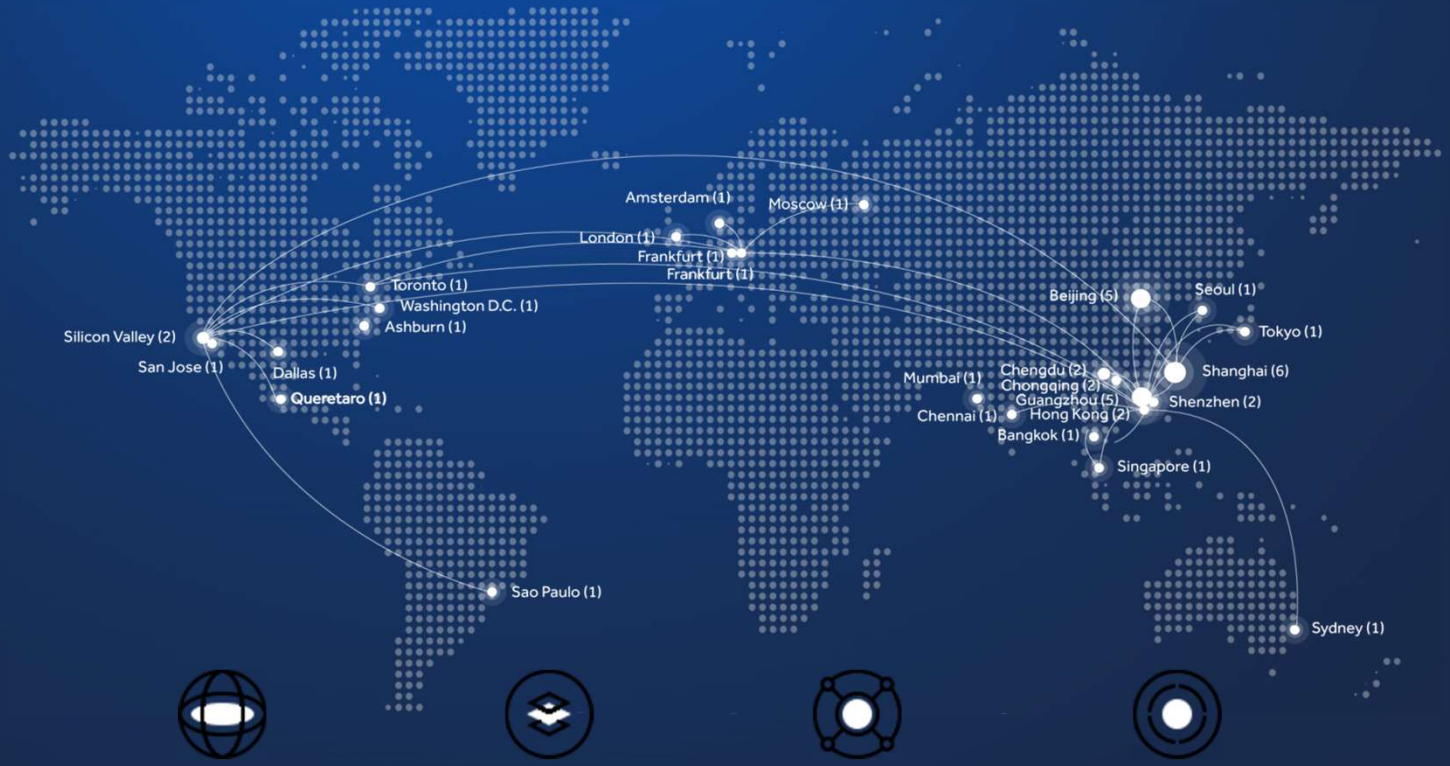
25 Regions

53 Availability zones

1,100+ PoP

1,000,000+ Servers

1,024+ PB Storage



Powerful Network | Highly Customized | Global Coverage | Ecosystem

Tencent Cloud can serve globally with large scale of resources

Agenda

- Background
- Design
- Future works

Tencent Cloud

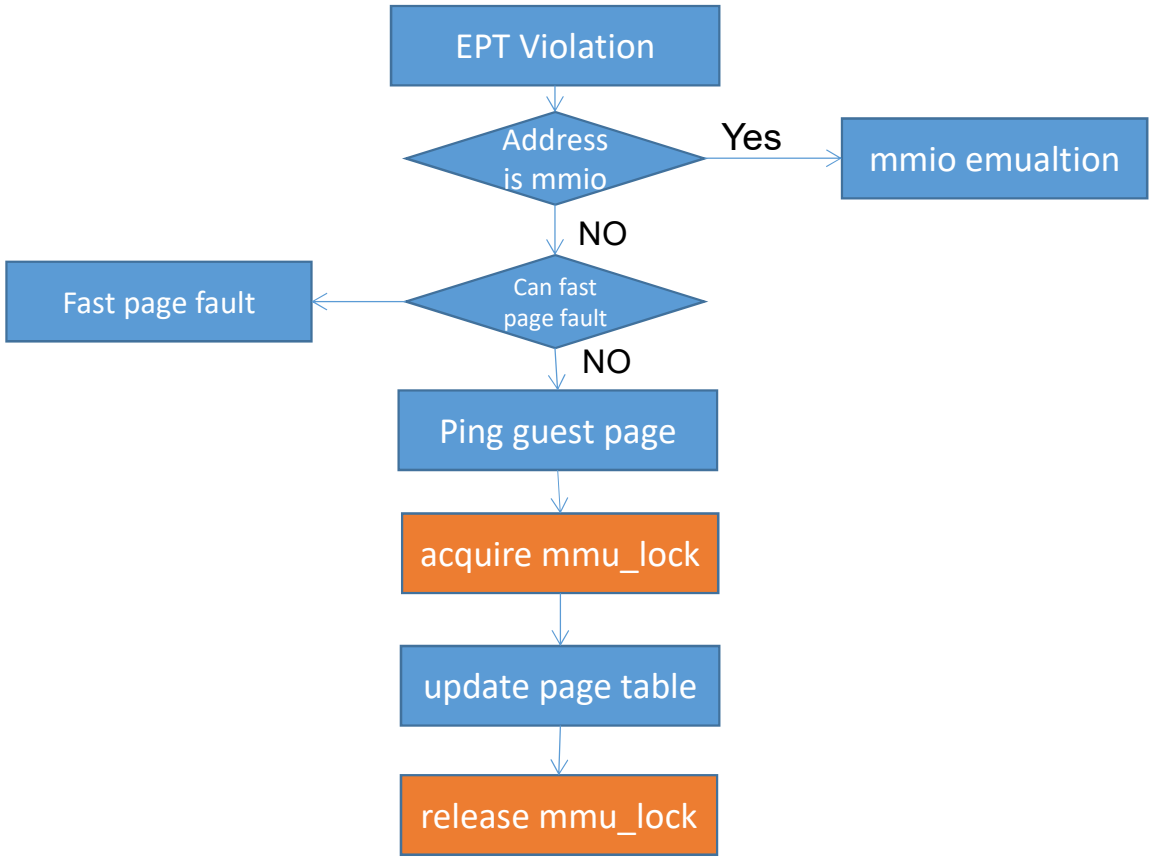
Background

- Significant performance drop after live migration
  - Multiple VCPUS
  - Numerous memory
  - with huge page table

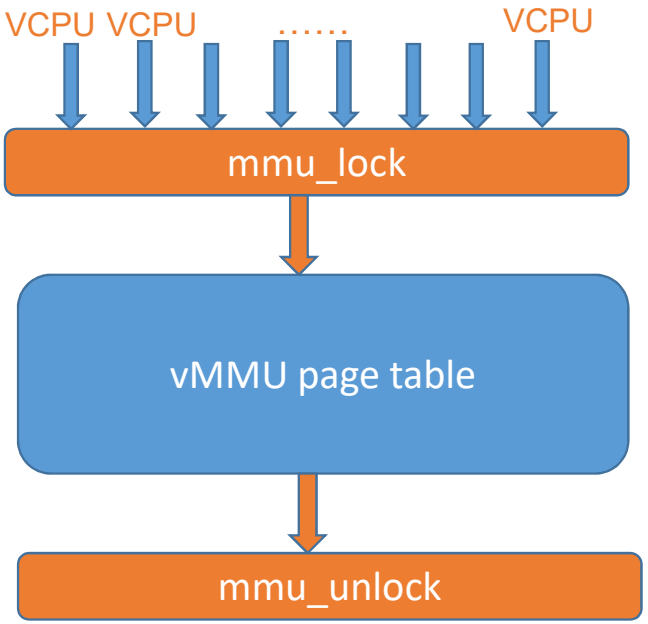
Overhead Analysis

- Numerous page fault happens after migration
- setup the page mapping for guest try to access memory
- enlarge the memory will cause performance decline

Current EPT setup



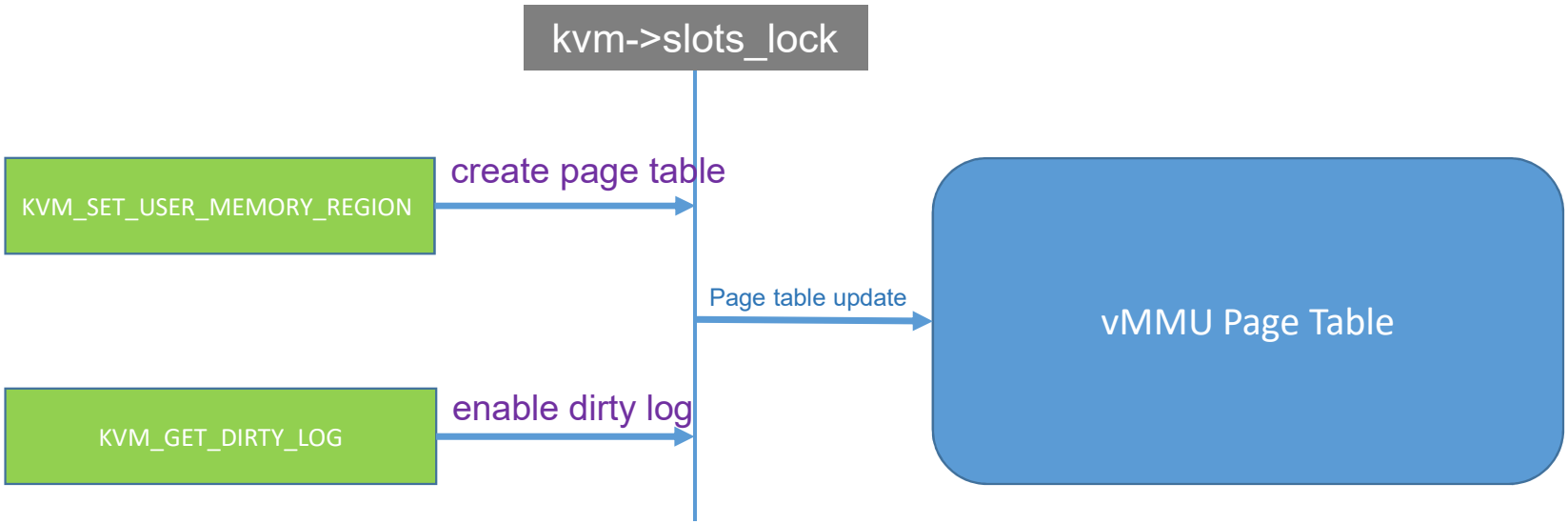
Bottleneck



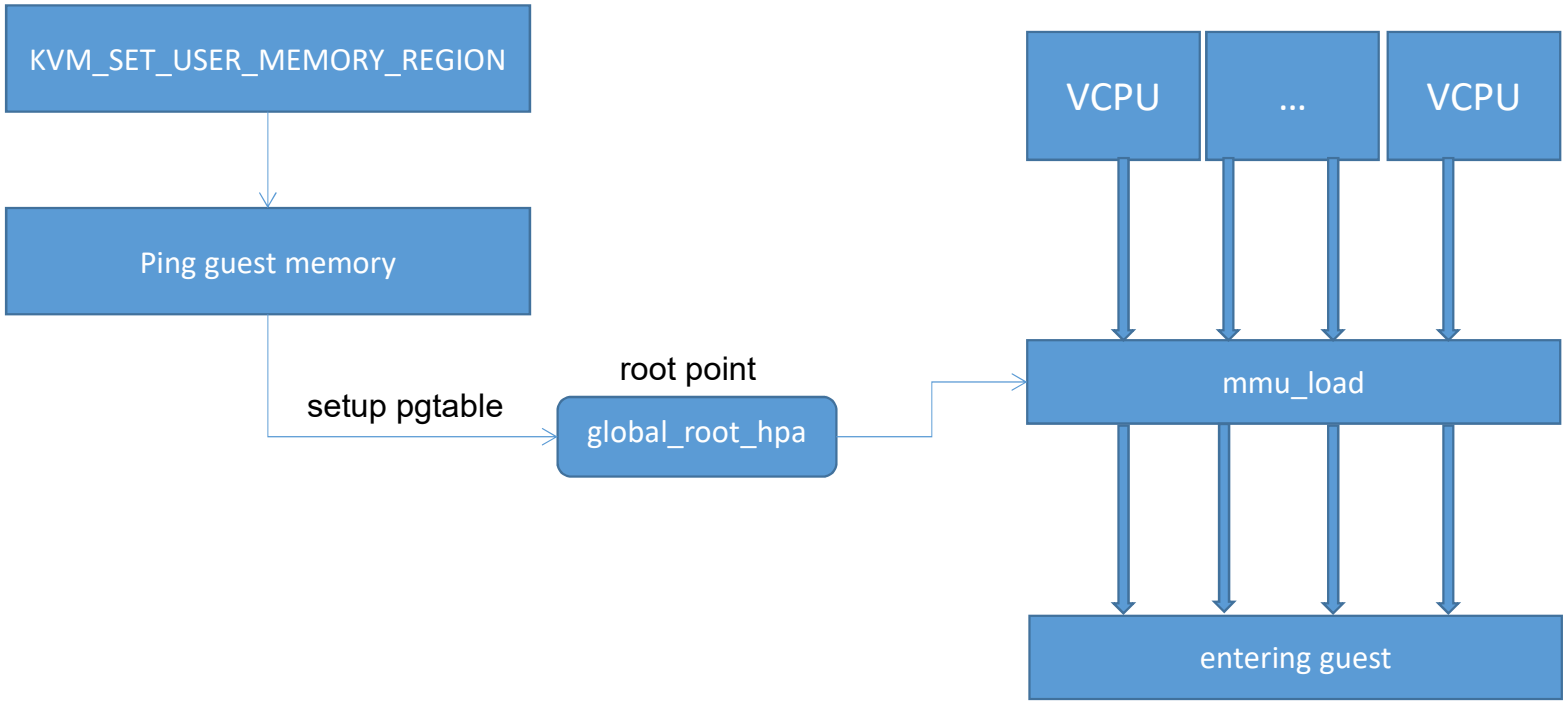
Our Proposal

- Improve concurrency of the vcpus with pre-setup page table
- lockless update the R/W status

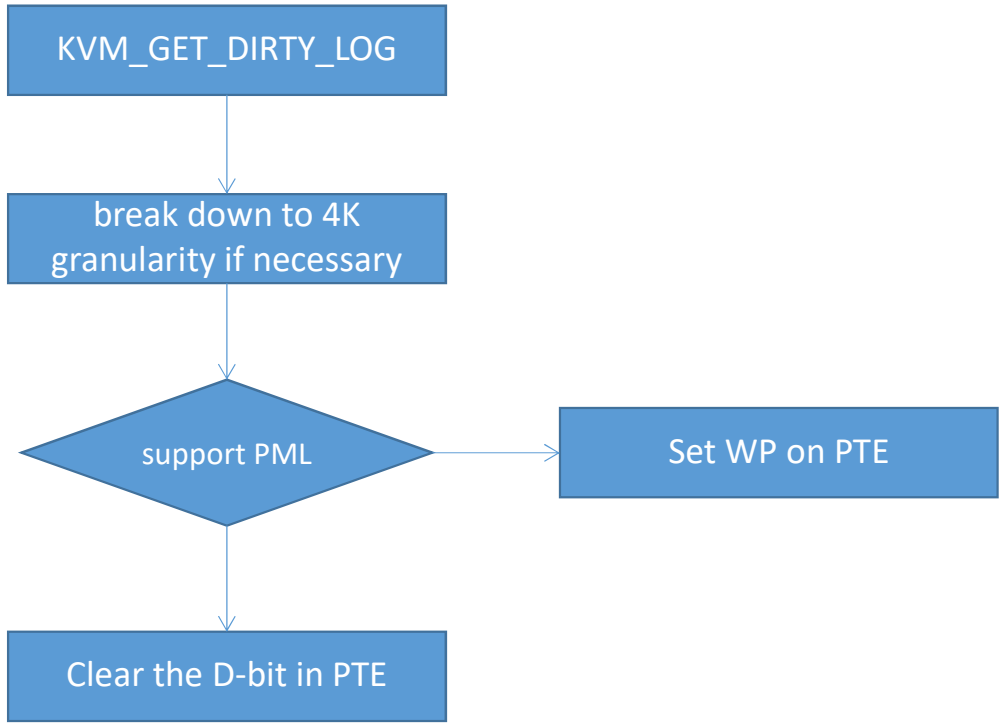
Overview


kvm->slots_lock
KVM_SET_USER_MEMORY_REGION
create page table
Page table update
vMMU Page Table
KVM_GET_DIRTY_LOG
enable dirty log

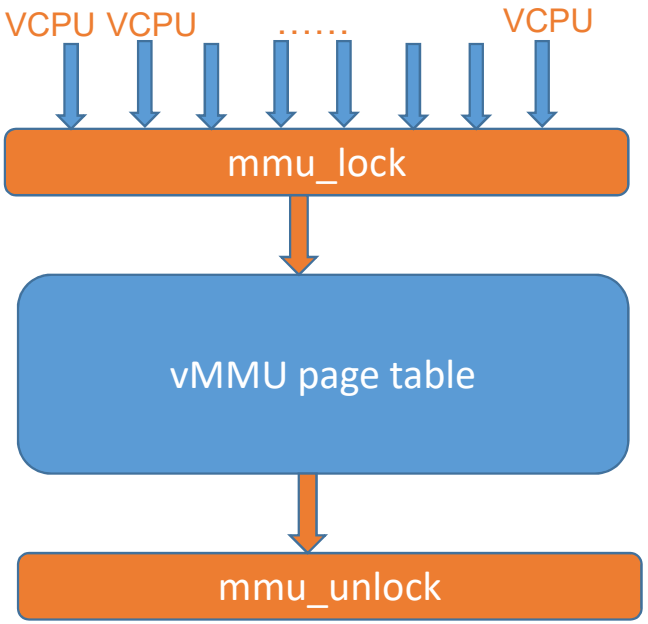
Fast pin guest memory and setup page table



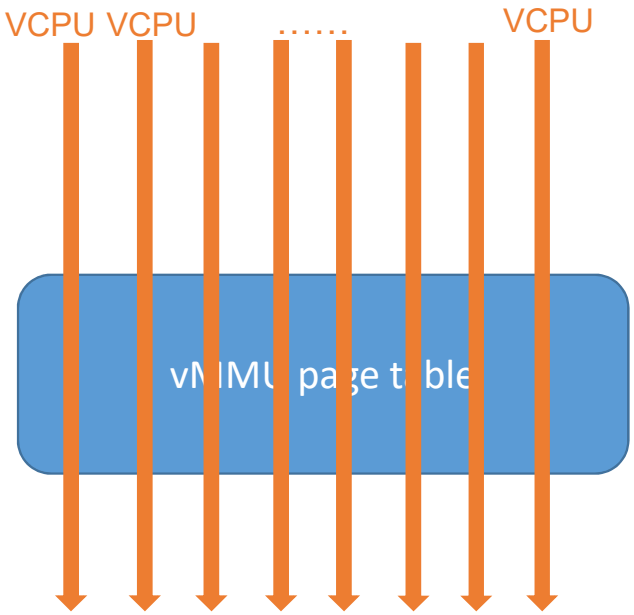
Live migration support



Update page entry with mmu_lock



Lockless access



Performance data

Test VM with 32 vCPUs and 64G memories, force each vCPU to dirty 2G memory.

Page size	Normal (second)	Pre-population(second)
4K	18~21	2~2.5
2M	3.2~3.6	2~2.5

Benefit for Parallel Memory Virtualization

- lockless access
- Eliminate the page fault latency
- Save system resource
  - mmu notification
  - Shadow page caches
  - Parent reverse mapping

Limitation

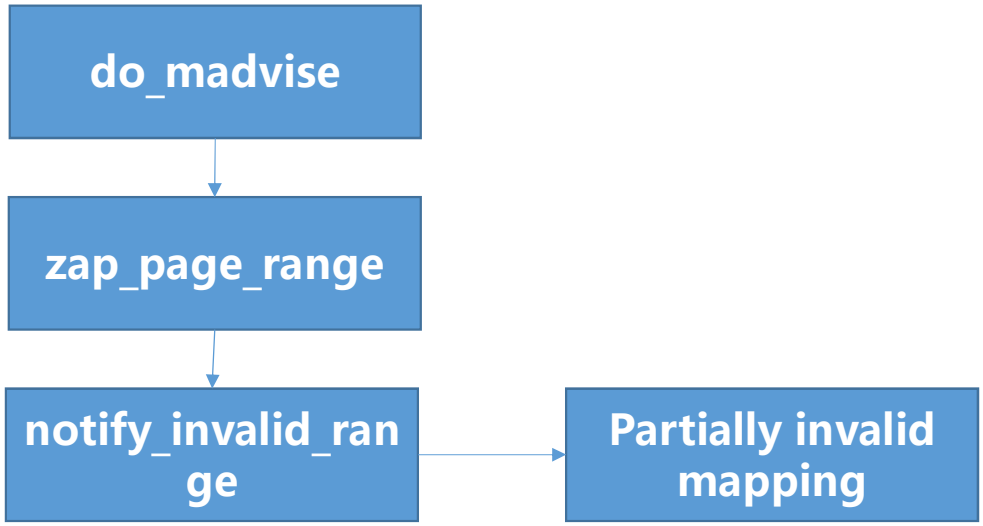
1. TDP enabled mode
2. SMM is not supported
3. memory overcommit is not supported

Future works

Tencent Cloud

Future works

- Post copy in live migration



Link to the source code

https://lkml.org/lkml/2020/9/1/425

Q&A

Tencent Cloud