VMware ESX Server Memory Resource Management

Aidan Lloyd-Tucker

Williams College

May 8, 2022



► Hypervisor





HypervisorPart of vSphere





- Hypervisor
- Part of vSphere
- Widespread industry use



VMware ESXi

- Hypervisor
- Part of vSphere
- Widespread industry use
- In tons of data centers



VMware ESXi

- Hypervisor
- Part of vSphere
- Widespread industry use
- ► In tons of data centers
- ► Bare metal (now)



VMware ESXi

- Hypervisor
- Part of vSphere
- Widespread industry use
- In tons of data centers
- ► Bare metal (now)
- Runs unmodified OSes





 Resource management policies compute target memory allocation for each VM

- Resource management policies compute target memory allocation for each VM
- Allocations are enforced via low-level mechanisms to reclaim memory from VMs

- Resource management policies compute target memory allocation for each VM
- Allocations are enforced via low-level mechanisms to reclaim memory from VMs
- Background activity exploits opportunities to share identical pages between VMs

► *Machine address* — hardware

- ► *Machine address* hardware
- Physical address virtualized memory that looks pysical to the guest

- ► Machine address hardware
- Physical address virtualized memory that looks pysical to the guest
- pmap physical to machine

- ► Machine address hardware
- Physical address virtualized memory that looks pysical to the guest
- pmap physical to machine
- shadow page tables guest OS virtual to machine

- Machine address hardware
- Physical address virtualized memory that looks pysical to the guest
- pmap physical to machine
- shadow page tables guest OS virtual to machine



Balloning

Coaxes the guest OS into using less memory

Figure 6: Inflating the balloon in a virtual machine ESX



Balloning

- Coaxes the guest OS into using less memory
- ► Transfers pressure from host to guest OS

Figure 6: Inflating the balloon in a virtual machine ESX



Balloning

- Coaxes the guest OS into using less memory
- Transfers pressure from host to guest OS
- ESX Server then reclaims ballooned memory



Figure 6: Inflating the balloon in a virtual machine ESX

Eliminates redundant copies of pages (code, read-only data) like COW

- Eliminates redundant copies of pages (code, read-only data) like COW
- No change to guest OS code

- Eliminates redundant copies of pages (code, read-only data) like COW
- No change to guest OS code
- Identify page copies by their contents via hashing

- Eliminates redundant copies of pages (code, read-only data) like COW
- No change to guest OS code
- Identify page copies by their contents via hashing



Figure 4. Content-Based Page Sharing in ESXi

Traditional OS adjusts memory allocations to improve performance

- ► Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities

- Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities
- Better to penalize less important VMs even if it has the highest marginal benefit of more memory

- Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities
- Better to penalize less important VMs even if it has the highest marginal benefit of more memory
- Charge the client more for an unused idle page

- Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities
- Better to penalize less important VMs even if it has the highest marginal benefit of more memory
- Charge the client more for an unused idle page
- Pages reclaimed preferentially from VMs not actively using their full allocation

- Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities
- Better to penalize less important VMs even if it has the highest marginal benefit of more memory
- Charge the client more for an unused idle page
- Pages reclaimed preferentially from VMs not actively using their full allocation
- Tax rate = maximum fraction of idle pages that can be reclaimed

- Traditional OS adjusts memory allocations to improve performance
- In vSphere, each VM has different allocations and priorities
- Better to penalize less important VMs even if it has the highest marginal benefit of more memory
- Charge the client more for an unused idle page
- Pages reclaimed preferentially from VMs not actively using their full allocation
- Tax rate = maximum fraction of idle pages that can be reclaimed
- Measurement = statistically sample pages in each VM