Virtual Machine

Memory Management I

資工三 B00902064 宋昊恩
資工三 B00902110 余孟桓

Outline

- Brief introduction
- Motivation
- Terminology
- Memory Demand Detection
- Memory Reclamation
- Memory Sharing
- Memory Utilization and Isolation
- Conclusion

Brief Introduction

Brief introduction

- Memory Resource Management in VMWare ESX server [1]
  - C. A. Waldspurger, OSDI '02: Proceedings of the 5th symposium on Operating systems design and implementation, 2002
- Collaborative Memory Management in Hosted Linux Environments [2]
  - Martin Schwidefsky, et. al., Linux Symposium, 2006
- Satori: Enlighted Page Sharing [3]
  - G. Milos, et. al., USENIX 2009
- Dynamic memory balancing for virtual machines [4]
  - Weiming Zhao, et. al., ACM SIGOPS Operating Systems Review, Volume 43 Issue 3, July 2009

Motivation

Motivation

- Two solution direction in terms of mechanism:
  - Memory Reclamation
  - Memory Sharing

- Two solution direction in terms of emphasis phase:
  - Memory Utilization
  - Memory Isolation

- Two solution direction in terms of fulfillment:
  - Modification on guest OS or not

Terminology

Terminology

- Overcommitment
  - Total size configured for all running VM exceeds the total amount of actual machine memory.
- Shadow Page Table
  - A mechanism for preserving the relationship between virtual address and machine address.
  - TLB then cache the mapping from Shadow Page Table.

Terminology (cont.)

- COW
  - Copy-on-write, a mechanism to save memory space. When a minor-refined copy is created, it is unnecessary to duplicate whole data, but only the differences between them.
- Paravirtualization (enlightenment, used in [3])
  - Different from full-virtualization, each guest OS system is aware of one another. In this case, the entire system can work together as a cohesive unit.

Memory Demand Detection

Memory Demand Detection

- Motivation:
  - To prevent VM idle and resources waste, it is important to get clear memory demand for each VM.

- Solution:
  - Sampling VM for specific span of time. [1]
  - LRU-based statistics. [4]

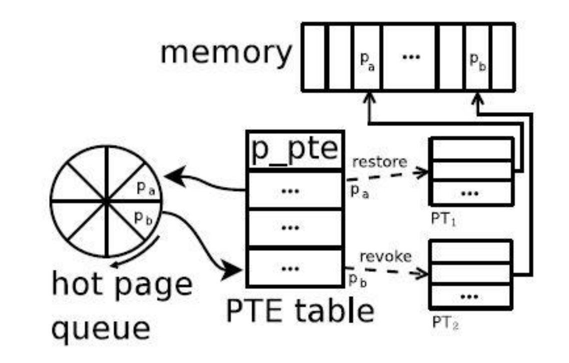
Sampling VM for specific span of time

- Description:
  - Determine span of time T, then invalidate N physical pages related TLB and MMU state in uniform distribution.
  - Once guest trying to re-establish mappings, counter increases.
- Drawbacks:
  - Tradeoff between estimation accuracy and overhead.
  - Cannot tell the VM performance when more or less memory is allocated.

LRU-based statistics

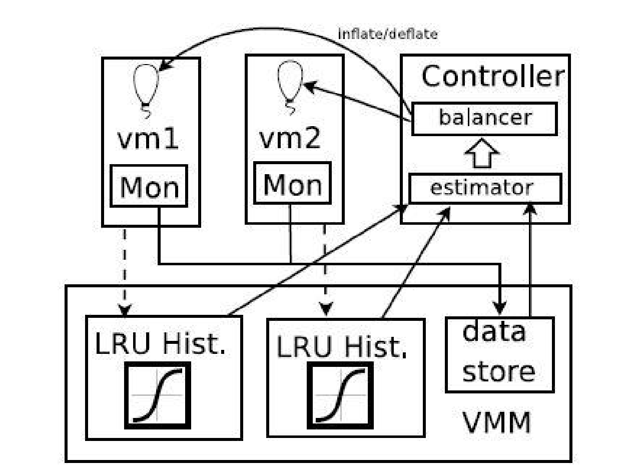
- Description:
  - Every page are cold initially, all accessibilities are removed.
  - When traps, it turns hot, accessibility is recovered.
  - When there are too many hot pages, LRU one will turn cold.
  - Trap only cold one, will not cause too much overhead.

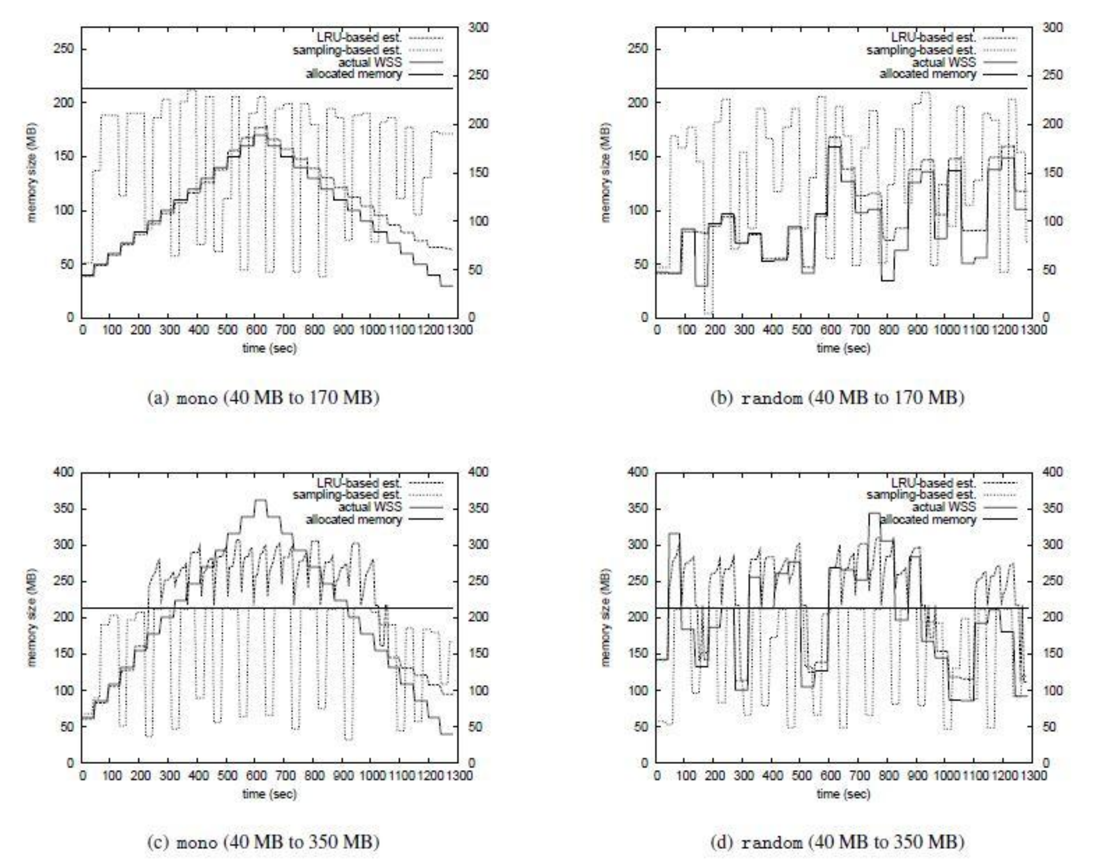
- Drawbacks:
  - Cannot detect swap usage.
  - Too many pages to preserve.
  - Updating needs linear search.



LRU-based statistics (cont.)

- Cannot detect swap usage
  - Add another background process to collect information, and send back to VMM.
- Too many pages to preserve
  - Collect G consecutive pages as a node.
- Updating needs linear search
  - Assume good locality exists.





(a) mono (40 MB to 170 MB)

(b) random (40 MB to 170 MB)

(c) mono (40 MB to 350 MB)

(d) random (40 MB to 350 MB)

Memory Reclamation

Memory Reclamation

- Motivation:
    - Commodity operating system don’t support dynamic changes to physical memory sizes.
    - But we still want to improve memory utilities.

- Solutions:
    - Page Replacement
    - Improved Page Replacement [1]
    - Ballooning [1]
    - MEmory Balancer (MEB) [4]

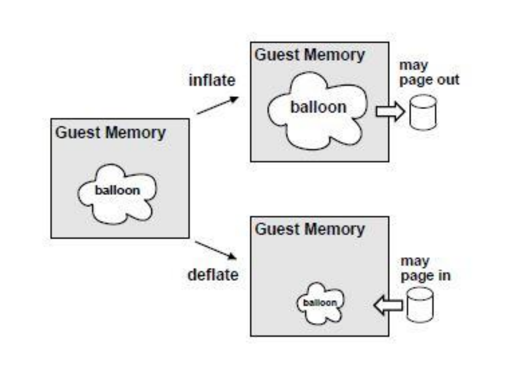
Page Replacement

- Description:
  - Host level pages
- Drawbacks:
  - Host-level knows little about guest OS memory situation.
  - Double Paging Problem
    - Page swapped out by host-level before swapping out by guest OS.
    - It can be improved by Randomized page replacement strategy (used by ESX server).

Ballooning

- Description:
  - A module loaded into guest OS as a pseudo-device driver.
  - When facing memory pressure, balloon inflates and notify VMM to reclaim physical pages it gain.

- Drawbacks:
  - It may be uninstalled, disabled explicitly.
  - It is not available while OS boosting.
  - Cannot reclaim memory quickly.
  - Each balloon has a minimum allocation.



Collaborative memory management (CMM)

- Collaborative memory management (CMM)
  - Infrequent Ballooning (CMM1)
    - Apply sufficient long term request
  - Page Replacement (CMM2)
    - Guest OS maintain *page status*
    - Host OS maintain *page resident*
  - Used by zSeries and z/VM

MEmory Balancer (MEB)

- Description:
  - Based on LRU statistics, it do dynamic memory resizing.
  - If the sum of VM memory demand can be fulfilled, then done.
  - Else, a delicate optimized algorithm is needed.

Memory Sharing

Memory Sharing

- Motivation:
  - Besides adjusting memory allocation between VMs, many VM shares similar process on same OS platform.
  - Prevent malicious virtual machine take advantage of memory reclamation.

- Solutions:
  - Transparent Page Sharing
  - Content-based Page sharing [1]
  - Enlightened page sharing [3]

Transparent Page Sharing

- Description:
  - Introduced by Disco, one copies are identified, multiple guest physical pages are mapped to the same machine page.

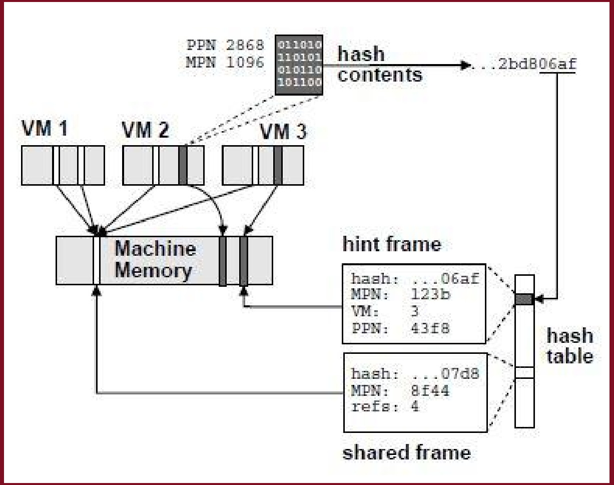
- Drawbacks:
  - It needs several guest OS modifications to identify copies.
    - For example, it hooked some code to routine “bcopy”.

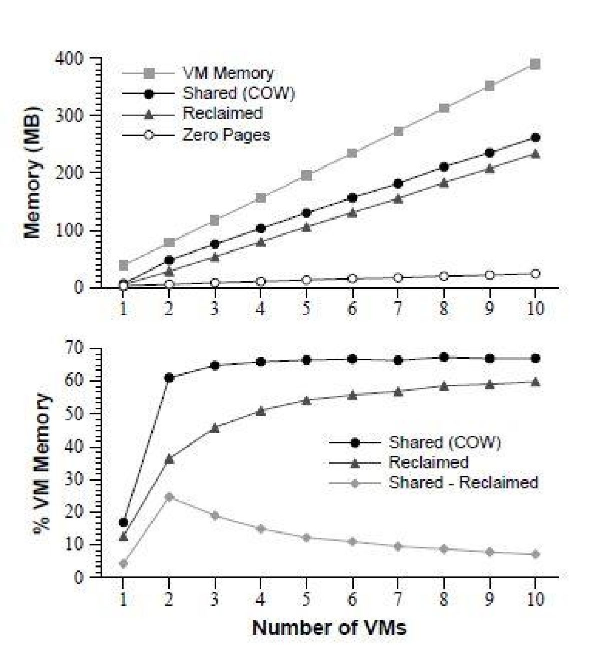
Content-based Page Sharing

- Description:
  - Initially, all pages are ordinary pages, and are hashed.
  - Once hash value meets, recalculate hash value.
    - If changed, remove old page and hash value.
    - If not changed, mark the page as COW.
  - It need to scan pages frequently.

- Drawbacks:
  - Tradeoff between scanning frequency and overhead.
  - Hard to discover short-lived share memory. (< 40 min)

Content-based Page Sharing (cont.)





Enlighted Page Sharing

- Description:
  - Modify virtual disk subsystem, and implement Sharing-aware Block Devices.
  - Detect sharing directly when data is read from disk.

- Advantages:
  - Avoid scanning overhead.
  - Detect short-lived sharing immediately.

Enlighted Page Sharing (Cont.)

- Drawbacks:
  - Cannot detect consequent memory writes.
  - Lots of modification, including hypervisor modification, sharing-aware block device addition, and adding repayment FIFO to guest OS kernel.

Memory Utilization and Isolation

Memory Utilization

- Motivation:
  - No matter how many resources one claim, the top principle is maximize the efficiency of memory usage.
- Advantages:
  - Avoid resources idle.
- Drawbacks:
  - Some malicious VM may utilize auto memory balancing mechanism to gain unreasonable amount of resources.

Memory Isolation

- Motivation:
  - No matter how imbalancing memory usage are, VM should follows isolation principles, and not affected by others.

- Advantages:
  - Can prevent malicious user, and preserve one’s privilege.

- Drawbacks:
  - Most of the time, malicious user does not exist.
  - Large amount of memory are wasted.

Some mechanisms

- Min-funding revocation:
  - One paid more money, the least valuable one becomes victim.

- Page-share advantage:
  - One share more pages with others uses more memory space.
  - No matter one page is shared or changed, only VMs sharing that certain page are involved.

Some mechanisms (cont.)

- Idle memory tax:
  - $\rho = S / (P \cdot (f + k \cdot (1 - f)))$
    - ρ : shares-per-page ratio
    - S : shares
    - P : number of pages
    - $f = t / n$: active rate
    - $k = 1 / (1 - \tau)$: idle page cost
    - τ : tax rate
    - t : touched page account
    - n : number of random sample pages

Conclusion

Conclusion

- There are always diverse choices for us to choose.
    - It is hard to balance between accuracy and overhead.
    - It is hard to balance between performance and isolation.
    - It is hard to determine whether revised OS is needed or not.