RAM Overcommit in Cloud

Yu Zhao

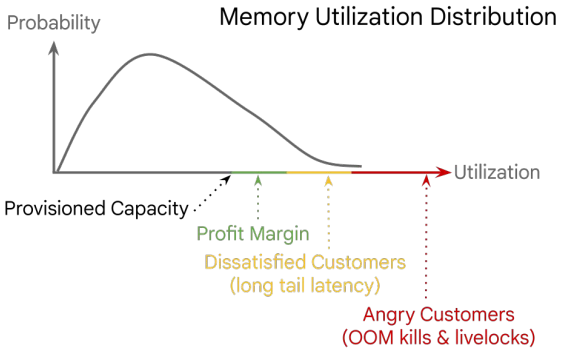
Why we care

Observations

- Applications' appetite for RAM is likely to keep growing
- RAM is likely to contribute more to the TCO in the future

Prediction

- RAM will continue to play a major role in profitability



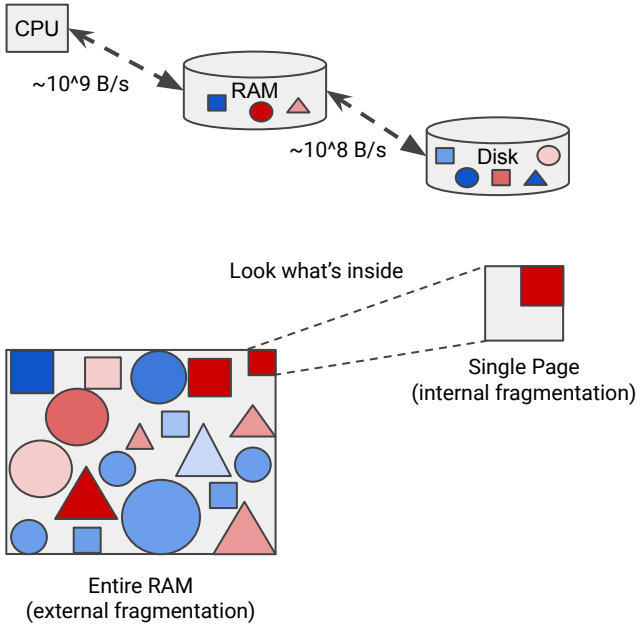
Fundamental problems

What objects to store in RAM

- Page replacement algorithms
- E.g., LRU

How to store more objects in RAM

- Defragmentations
- E.g., compression



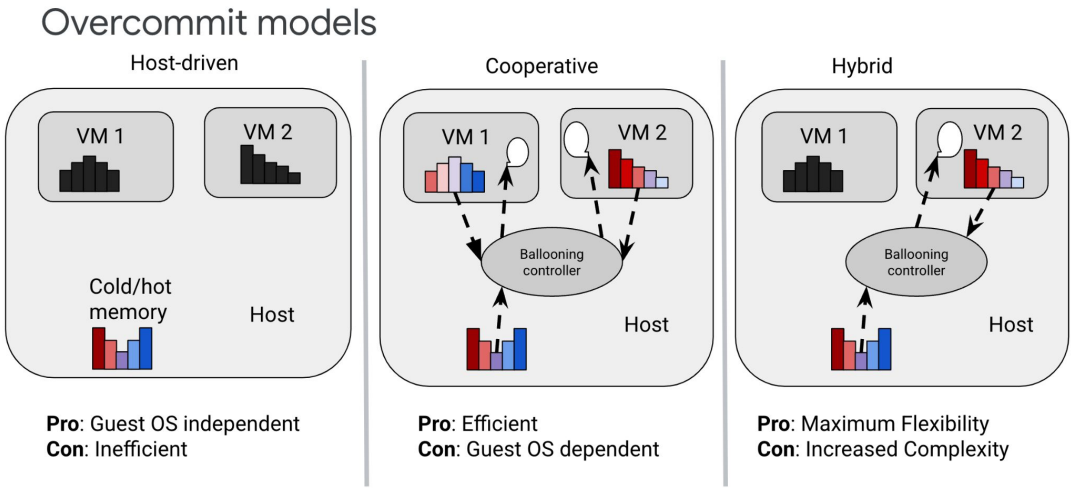
Building blocks

Userspace

- VM bin packing
- Working set size (WSS)

Kernel space

- Report WSS
- Reduce fragmentation



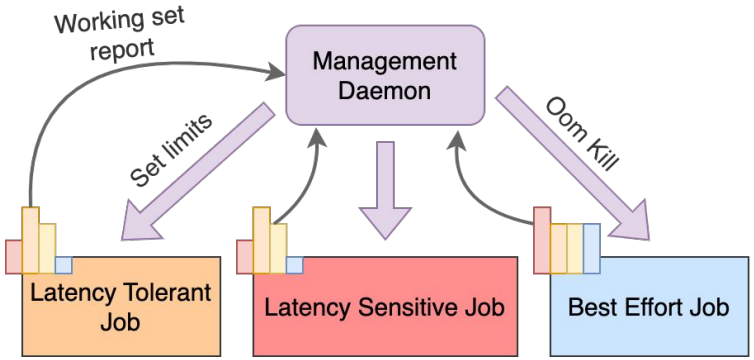
Working set size

WSS is a binning of pages

- Bin by measure of recency or frequency
- Can be from the host itself or its guests

Guest WSS gives a better estimate

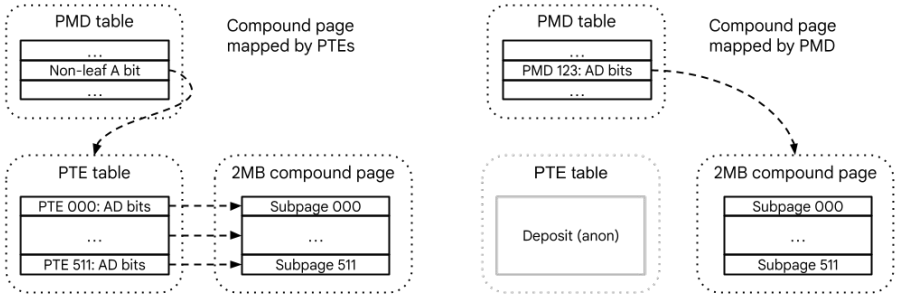
- Guests can cheat when reporting its WSS
- More suitable for sandboxing VMs, e.g., serverless



Fragmentations

Internal fragmentation

- Base page (4KB): compression
- Huge page (2MB): auto detection



External fragmentation

- Hosts: reduces ability to back VMs with 2MB THPs
- Guests: reduces ability to hot unplug 2MB page blocks

Policies and heuristics

When and how much

- Userspace: ML
- Kernel space: eBPF

One size fits ~~all~~ none!

Software-defined far memory in warehouse-scale computers
https://research.google/pubs/pub48551/

Additional considerations

Ballooning

- No double compression
- Can drop clean page cache in guest

Memory bandwidth

- Side effect from overcommit

Thank you!