Increasing Memory Density through Dynamic Memory Extension with Memory1 through Flash

Rellermeyer, Jan S.; Amer, Maher; Smutzer, Richard; Rajamani, Karthick

Publication date
2018

Document Version
Accepted author manuscript

Citation (APA)
Rellermeyer, J. S., Amer, M., Smutzer, R., & Rajamani, K. (2018). *Increasing Memory Density through Dynamic Memory Extension with Memory1 through Flash*. Poster session presented at ICT.OPEN 2018, Amersfoort, Netherlands.

Increasing Container Density through Dynamic Memory Extension with Memory1 Flash

Jan S. Rellermeyer*, Maher Amer†, Richard Smutzer†, Karthick Rajamani‡

*Distributed Systems Group, TU Delft, †Diablo Technologies, ‡IBM Research

Containers in Practice

- Containers co-exist on the same OS as opposed to full virtualization with a separate OS per tenant
- In practice, we see many cases where a majority of the containers on a server are mostly inactive for an extended period of time while few containers show high activity.
- Should allow for higher density and better server utilization if we can pack container more densely without compromising the performance of the critical workloads

Container Density Benchmark

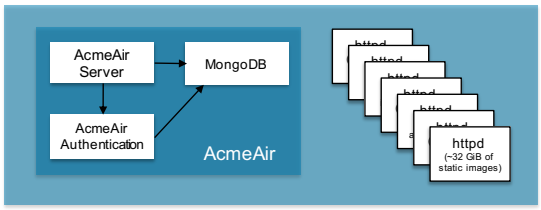

Figure 1: ContainerScale setup

One critical workload (AcmeAir) of 3 containers, varying the amount of mostly inactive *noise* containers (Apache httpd).

Evaluation

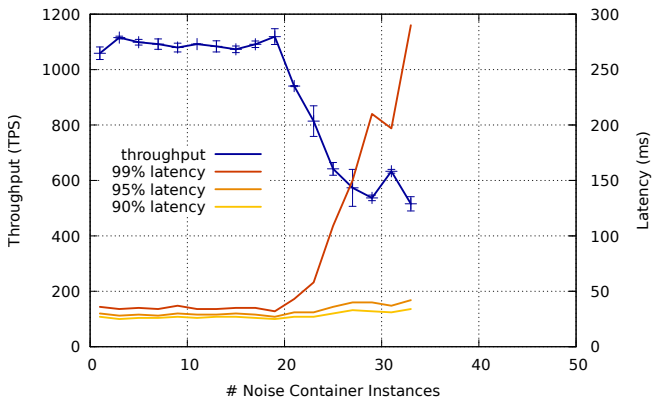

Figure 2: ContainerScale without DMX

Server[1] can only sustain around 19 noise containers despite an overall constant workload.

DOES NOT SCALE
Memory becomes the bottleneck and the tail latency of the critical workload explodes

Memory1 DMX

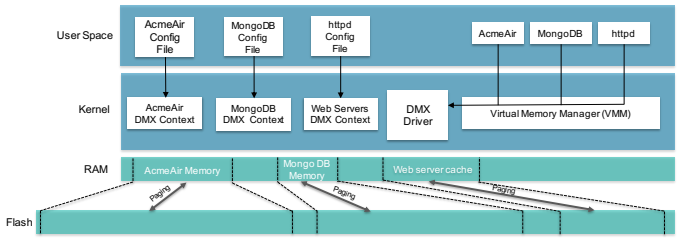

Figure 3: Memory1 DMX

Memory1 is a server memory extension product developed by Diablo Technologies that plugs into the DDR-4 memory channel and provides high bandwidth, low latency access to flash devices mounted on the module. The DMX kernel driver intercepts and services all memory requests generated by the selected application (malloc, page fault, etc). DMX creates a *Memory Context* for each selected application running on the server and carves out a *dynamic* portion of the server's physical memory (i.e. DRAM) to be used as front-end cache for that application.

Evaluation with DMX

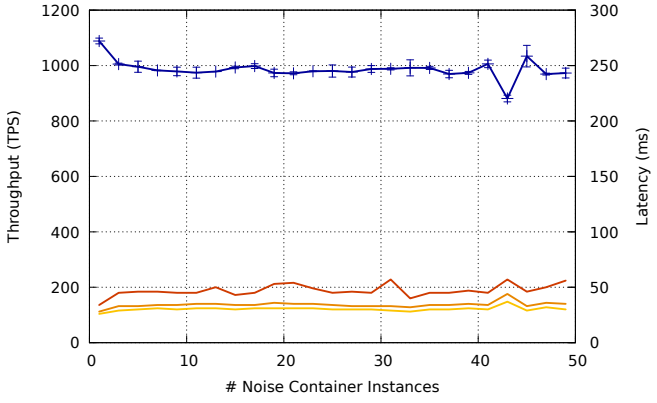

Figure 4: ContainerScale with DMX

The system can now sustain 50 noise container instances without significant performance degradation.

Dynamic memory extension with DRAM Flash helps to increase container density.

[1]Inspur NF5180M4 system with 2xIntel Xeon E5-2660 v3 (32 cores), 256GiB RAM, Linux 3.10 with Docker 17.03.1-ce, Docker OOM disabled. 2 TiB of Diablo Memory1 Flash