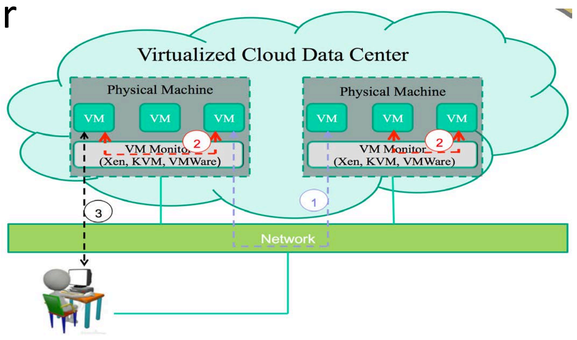
# Shared Memory Optimization in Virtualized Cloud

Qi Zhang and Ling Liu, Georgia Institute of Technology

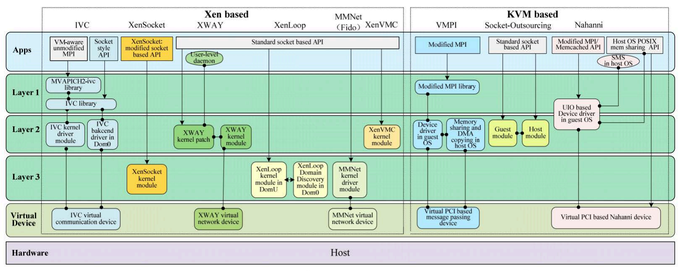
## Network I/O in virtualized cloud

Disadvantages of co-located inter VM communication via default network stack

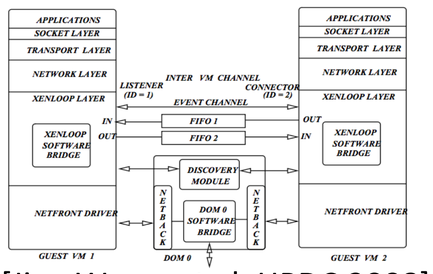
- Long communication path
- Multiple data copies
- Context switch between VMs and the host machine




## Current shared memory approaches




[Yi Ren, et al., CLOUD 2013]

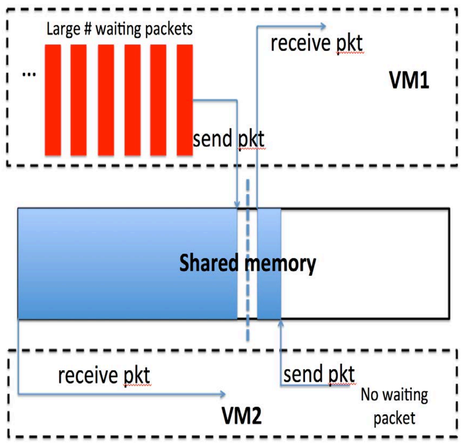



[Jian Wang, et al. HPDC 2008]

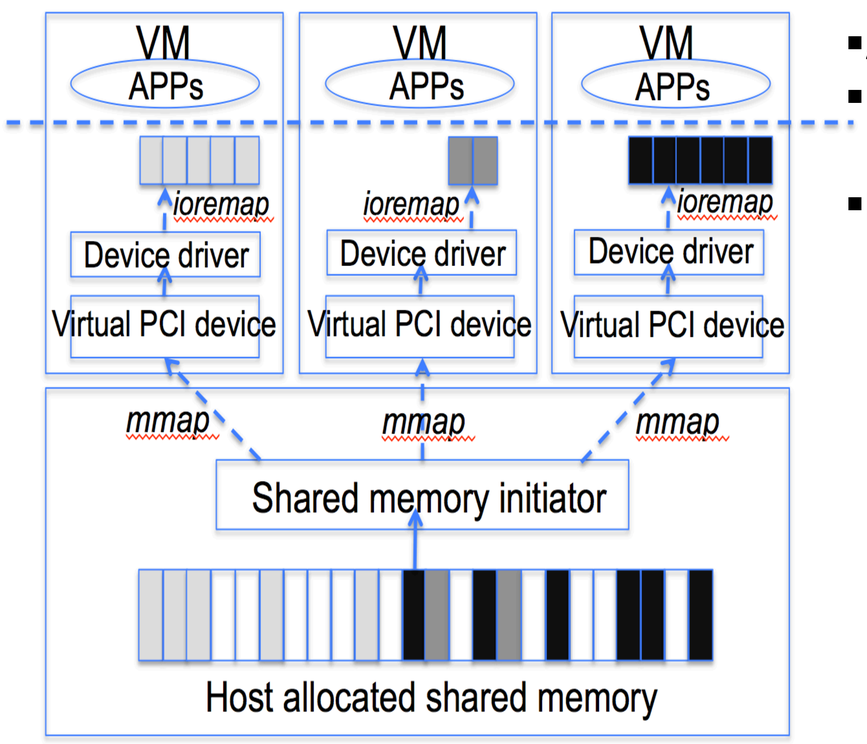
- Different layers where shared memory is established
  - User libraries and system calls level
  - Above the transportation level
  - Below the IP level

## Problems of current approaches

- Static shared memory allocation
  - Inflexibility
  - Low resource utilization
  - Restricted application performance
- Allocating shared memory from inside the VM
  - Increasing VM memory pressure
  - VM crash propagation



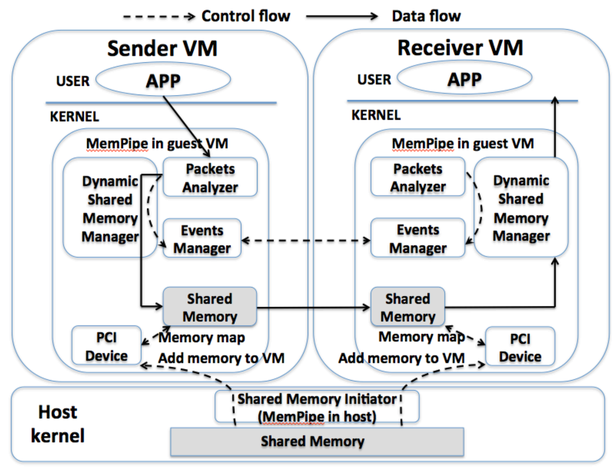

## Dynamic memory allocation from host




- Allocating shared memory from the host
- On-demand shared memory allocation to each VM
- Interfaces to VM kernel
  - struct shm* shn_init(size_t size)
  - void shm_exit(void *addr)
  - size_t shm_put(void *buf, size_t offset, size_t len)
  - Etc.

## Case study 1: MemPipe

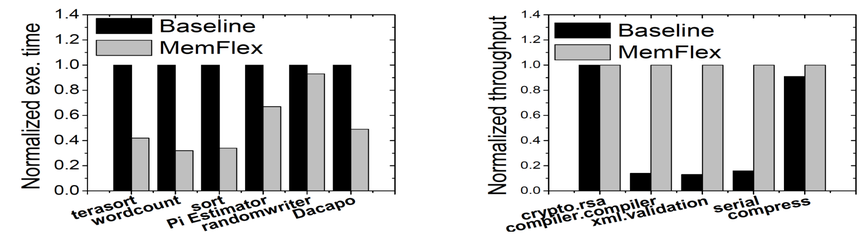
MemPipe is a co-located inter VM communication system built on top of the host machine based dynamic managed shared memory.




## MemPipe performance

COMPARISON OF SHARED MEMORY UTILIZATION, **udp sender**

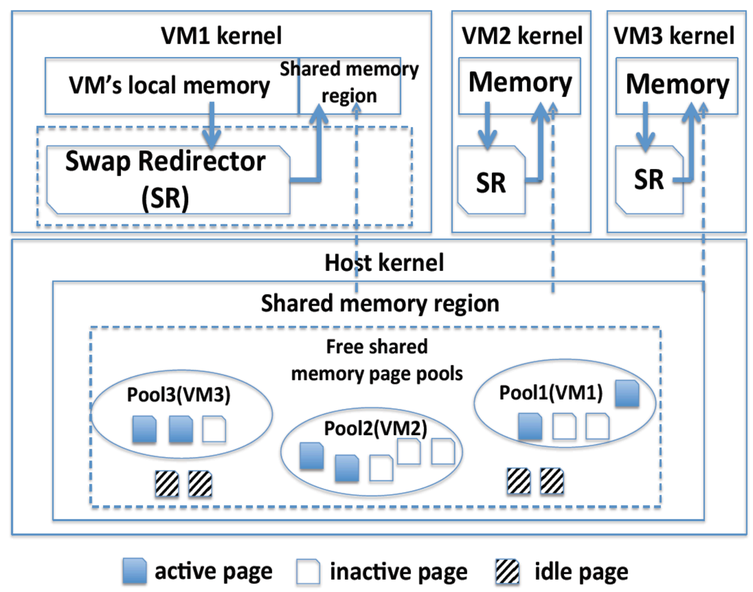
| msg size (Byte) | Static (Average values) | | | | Dynamic (Average values) | | | |
|---|---|---|---|---|---|---|---|---|
| | Allocated(KB) | Used(KB) | Util | Throughput (Mbps) | Allocated(KB) | Used(KB) | Util | Throughput (Mbps) |
| 64 | 1024 | 6 | **0.5%** | 128 | 14 | 6 | **45%** | 124 |
| 128 | 1024 | 11 | **1.0%** | 258 | 18 | 10 | **57%** | 246 |
| 256 | 1024 | 17 | **1.6%** | 503 | 26 | 17 | **66%** | 495 |
| 512 | 1024 | 31 | **3.0%** | 1005 | 42 | 31 | **73%** | 947 |
| 1024 | 1024 | 55 | **5.2%** | 1790 | 70 | 56 | **80%** | 1755 |
| 2048 | 1024 | 100 | **9.5%** | 2846 | 124 | 108 | **87%** | 2815 |
| 4096 | 1024 | 216 | **20.6%** | 4570 | 227 | 205 | **90%** | 4516 |
| 8192 | 1024 | 409 | **39.0%** | 5233 | 451 | 415 | **92%** | 5198 |
| 16384 | 1024 | 823 | **78.5%** | 6006 | 814 | 814 | **94%** | 5904 |



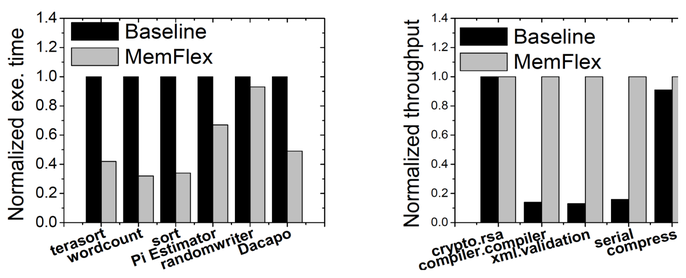

(a) Execution time(lower is better) (b) Throughput(higher is better)

## Case study 2: MemFlex

MemFlex is an efficient VM memory swapping system built on top of the host machine based dynamically managed shared memory.




## MemFlex performance




(a) Execution time(lower is better) (b) Throughput(higher is

Table 2: VM disk I/O performance measured by Sysbench

| S/R | Sequential | | Random | |
|---|---|---|---|---|
| R/W | Read (GB/s) | Write (KB/s) | Read (GB/s) | Write (KB/s) |
| Baseline | 3.89 | 19061.90 | 3.93 | 52.38 |
| MemFlex | 3.96 | **62444.54** | 4.17 | **1989.32** |

Intel Science & Technology Center for Cloud Computing

Carnegie Mellon University · Georgia Tech · intel · PRINCETON UNIVERSITY · UC Berkeley · UNIVERSITY of WASHINGTON