

# An Online Virtual Machine Placement Algorithm in an Over-Committed Cloud

Siqi Ji\*, Ming Da Li, Niannian Ji, Baochun Li



UNIVERSITY OF  
TORONTO



# Virtual Machine Placement

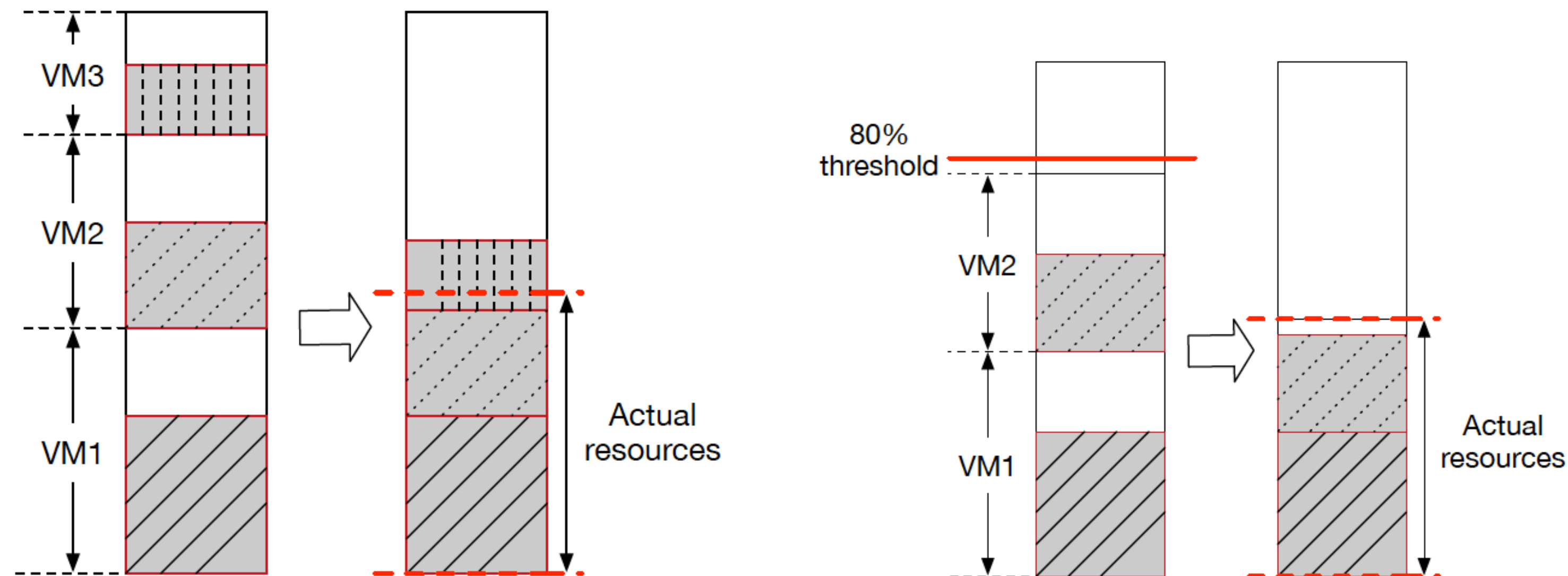
- ▶ Select the most suitable physical machine (PM) to host each virtual machine (VM).
- ▶ It is crucial to balance PM resources among multiple dimensions during the placement and minimize the number of activated PMs.

# Resource Over-Commitment

- ▶ Over-committed cloud: widely used for solving the wastage problem by allocating more resources to VMs than they actually have.
- ▶ Limitation of existing works:
  - ▶ Did not consider resource over-commitment in VM placement, which could cause PM overloading.

# PM Overloading

- ▶ Total resources utilized by VMs do exceed the PM's actual capacities.
- ▶ Memory of the PM is 36GB and it is sold as 72GB:



(a) Packing VMs as tightly as possible. (b) Setting a 80% threshold for placement.

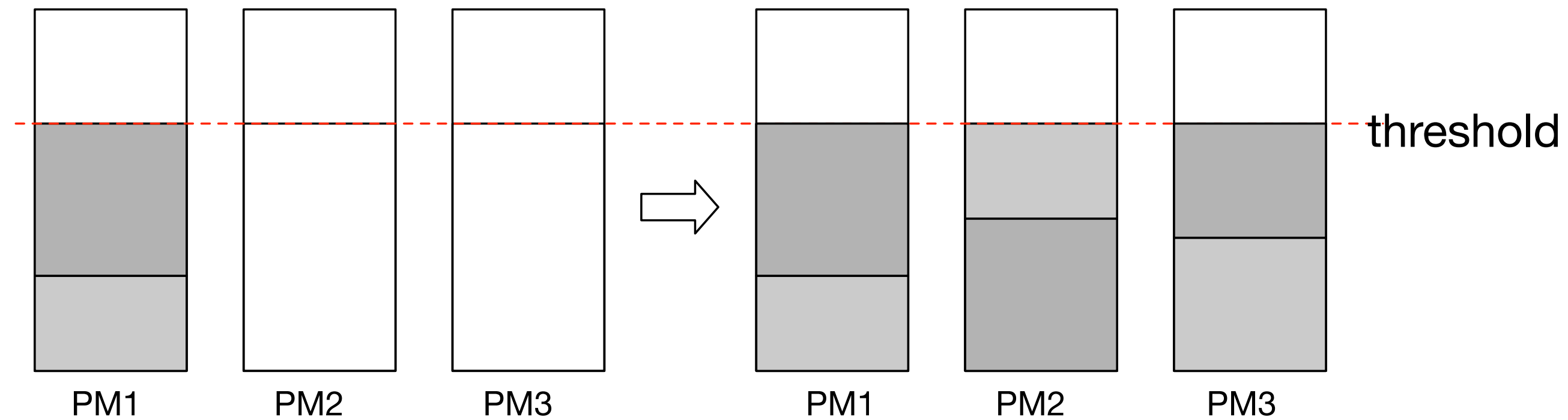
# Our Solution: Min-DIFF

- ▶ An threshold-based online VM placement algorithm that considers multiple dimensions of resources:
  - ▶ Reduce resource fragmentation
  - ▶ Reduce the risk of PM overloading

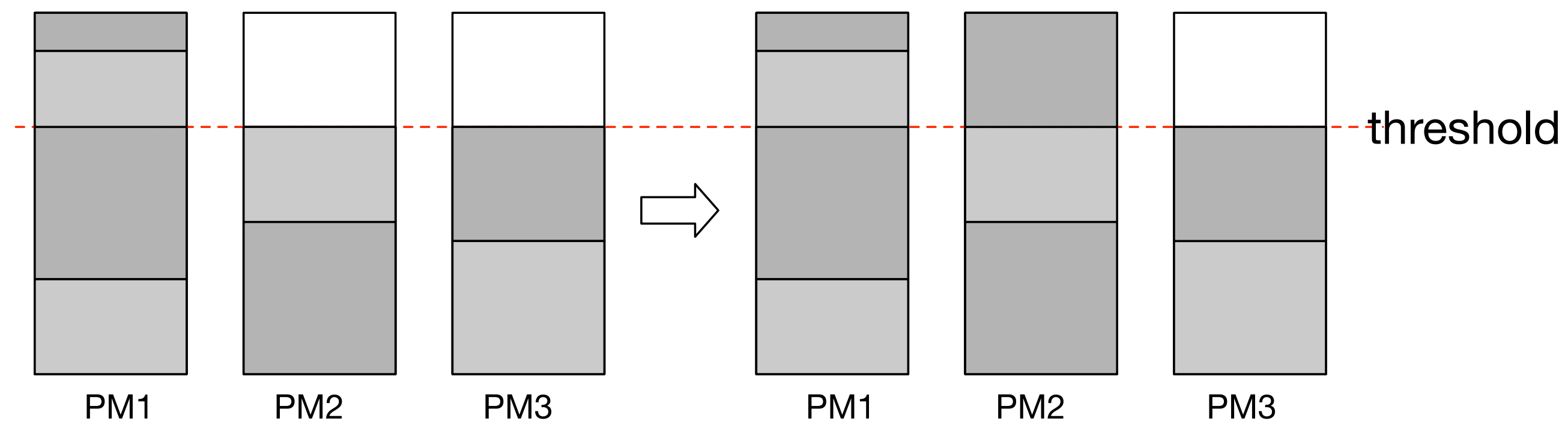
# Min-DIFF

## ► Threshold-based placement

Strategy 1: Place VMs below the threshold:



Strategy 2: Place VMs without considering the threshold



# Resource Threshold

- ▶ Warning line: providers do not expect the utilization of over-committed PMs is higher than a specific percentage.
- ▶ Reserve space for large VMs above the threshold.

$$Th_j^d = \min \left\{ \frac{PM_j^d - L^d}{PM_j^d}, w^d \right\}.$$

# Choose the Best PM

- ▶ Utilized PMs: Choose the PM that has the largest resource fragmentation reduction.

$$RF_j = \frac{\sum_{p,p \neq m} (NR_j^p - NR_j^m)}{\sum_{d=1}^D NU_j^d}, \quad \delta_{RF_j} = RF_{j\_bef} - RF_{j\_aft}.$$

- ▶ Empty PMs: Choose the most balanced PM after the VM is placed.

$$RF_{j\_empty} = \sum_{p,p \neq m} (NR_j^p - NR_j^m).$$



# Performance Evaluation

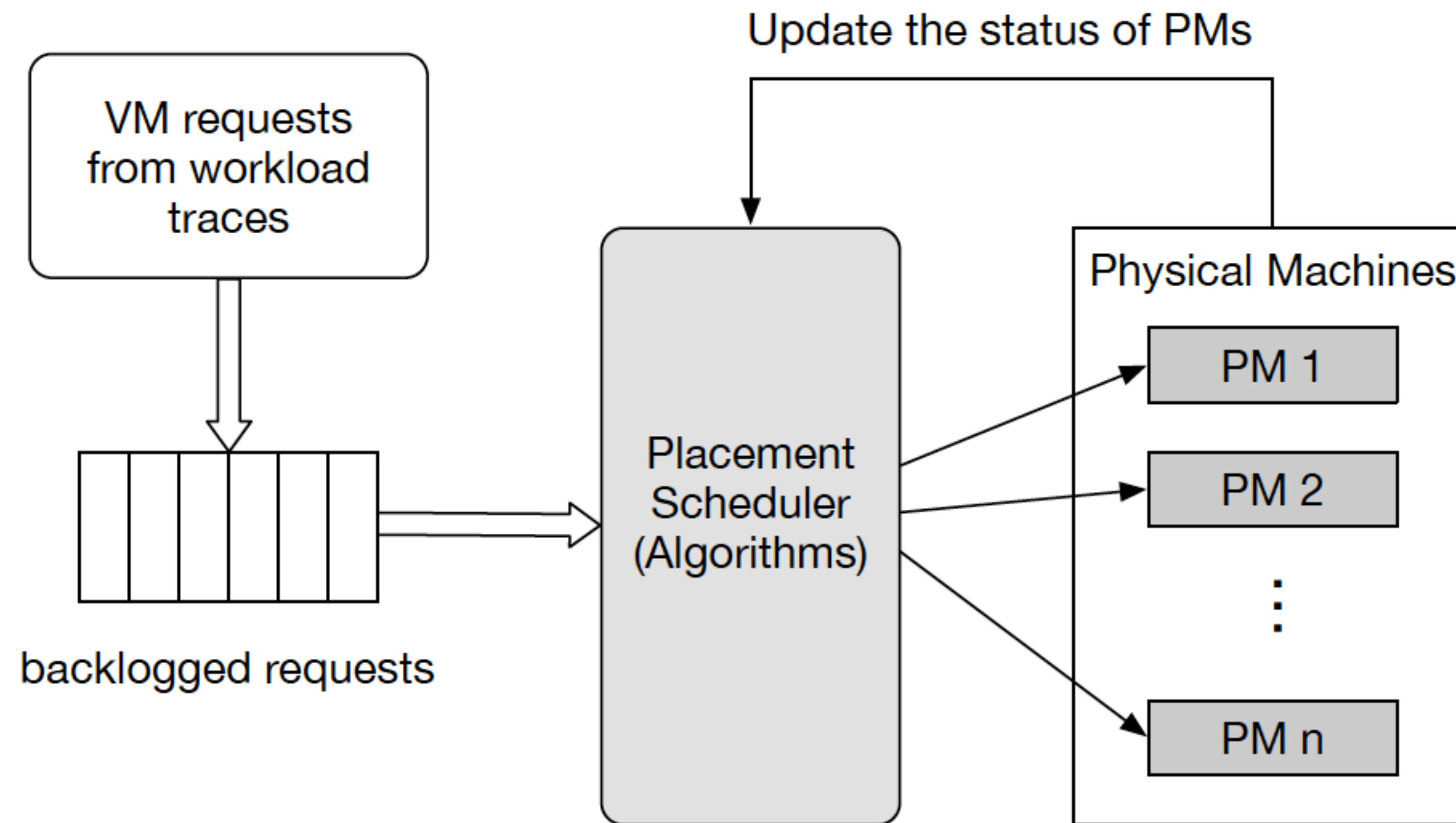
- ▶ Schemes for comparison:
  - ▶ First Fit algorithm
  - ▶ EAGLE [1]
  - ▶ Max-BRU algorithm [2]
- ▶ Three datasets we generated and one real-world workload Trace.

[1] X. Li, Z. Qian, S. Lu, and J. Wu, "Energy Efficient Virtual Machine Placement Algorithm with Balanced and Improved Resource Utilization in a Data Center," *Mathematical and Computer Modelling*, vol. 58, no. 5, pp. 1222–1235, 2013.

[2] N. T. Hieu, M. Di Francesco, and A. Y. Jaaski, "A Virtual Machine Placement Algorithm for Balanced Resource Utilization in Cloud Data Centers," in *Proc. IEEE International Conference on Cloud Computing (CLOUD)*, 2014.

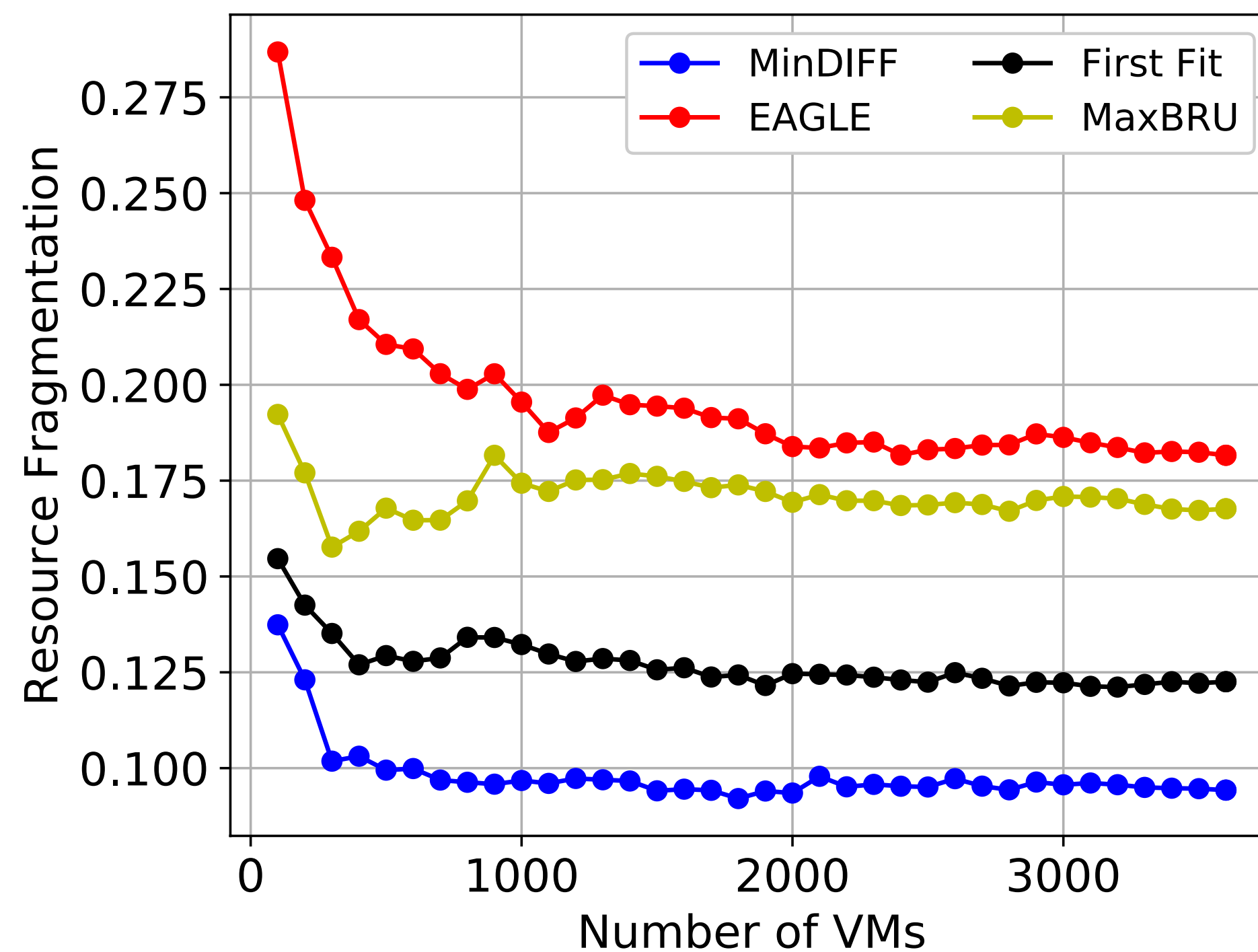
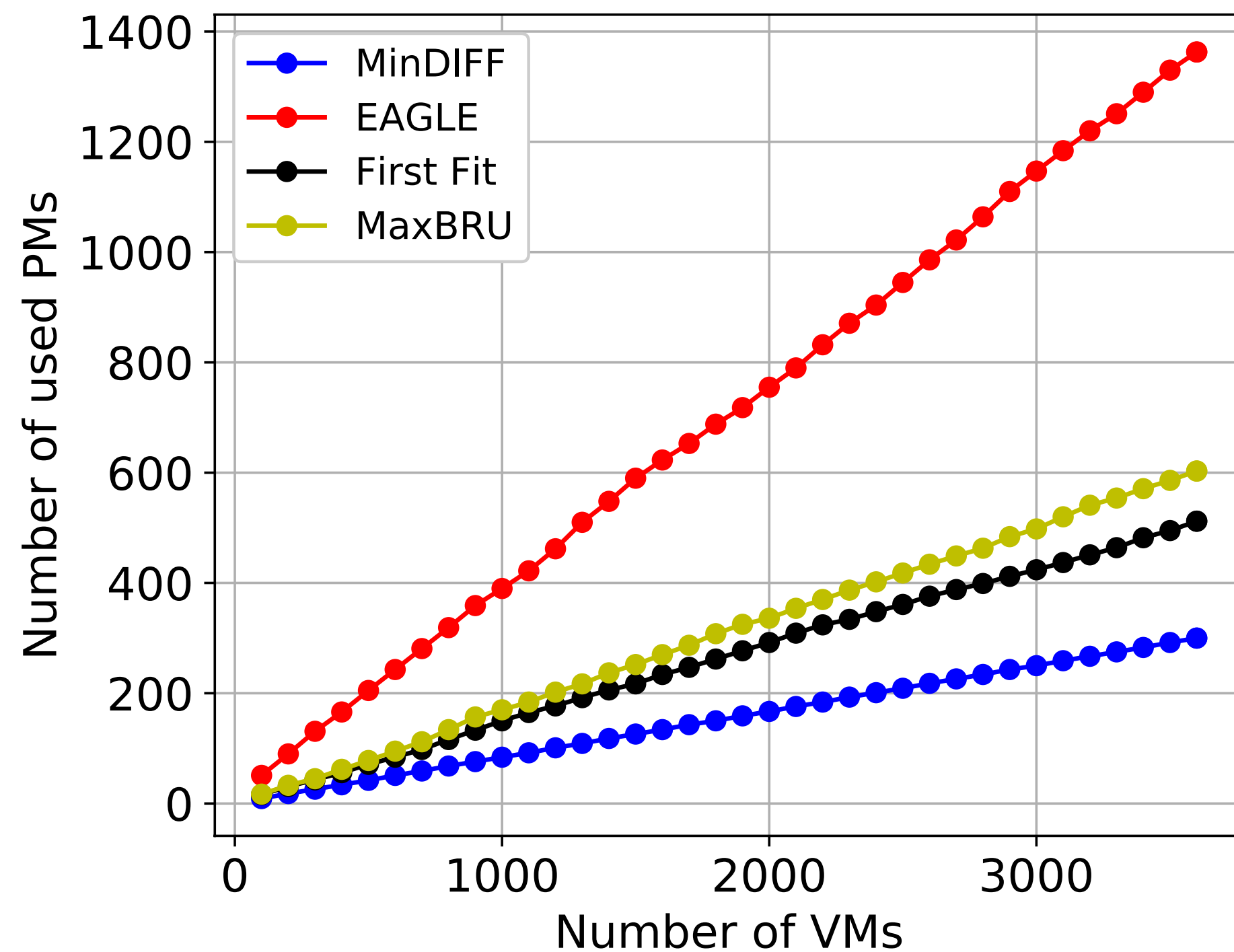
# Performance Evaluation

- ▶ Architecture of the simulator:



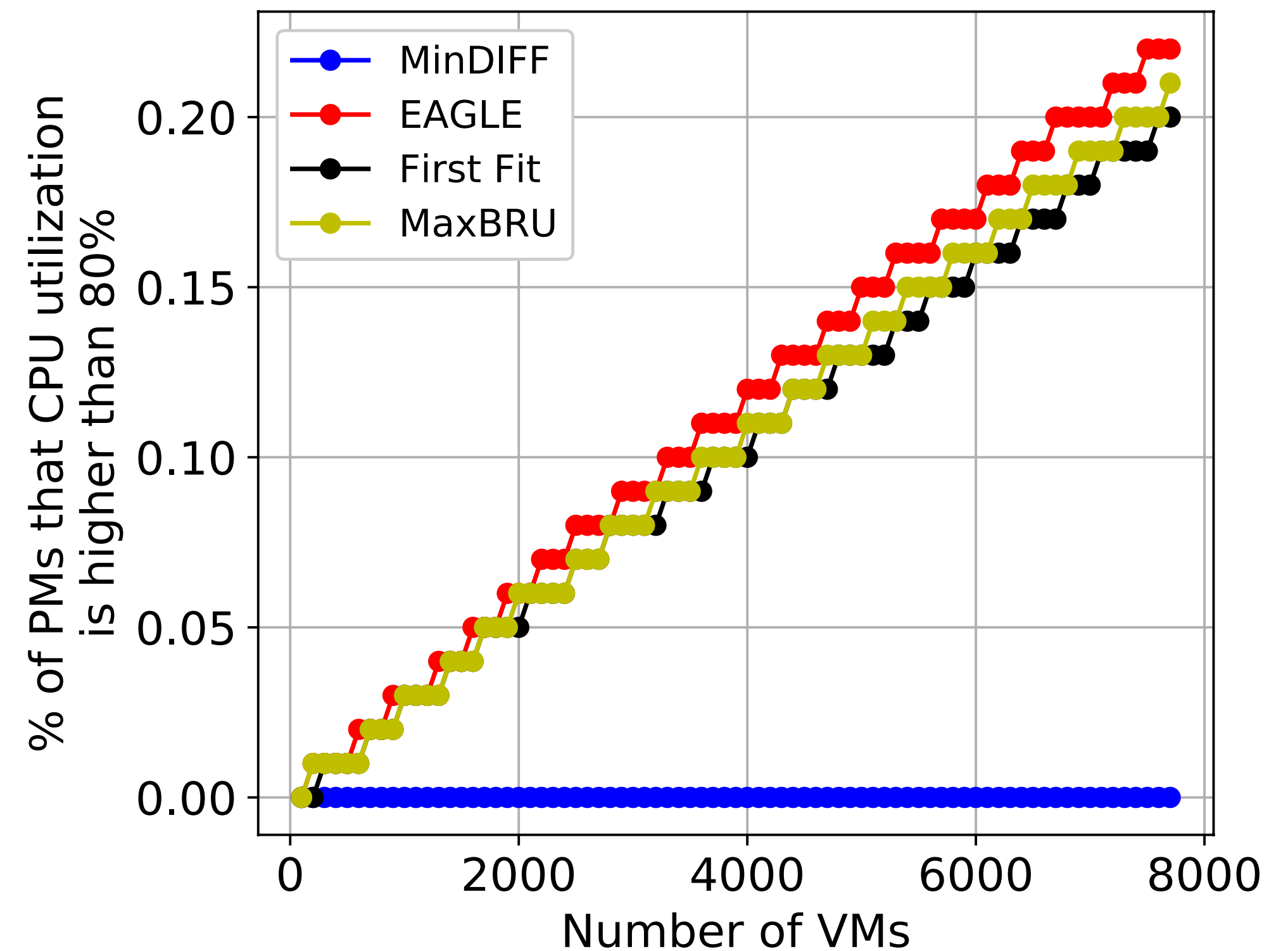
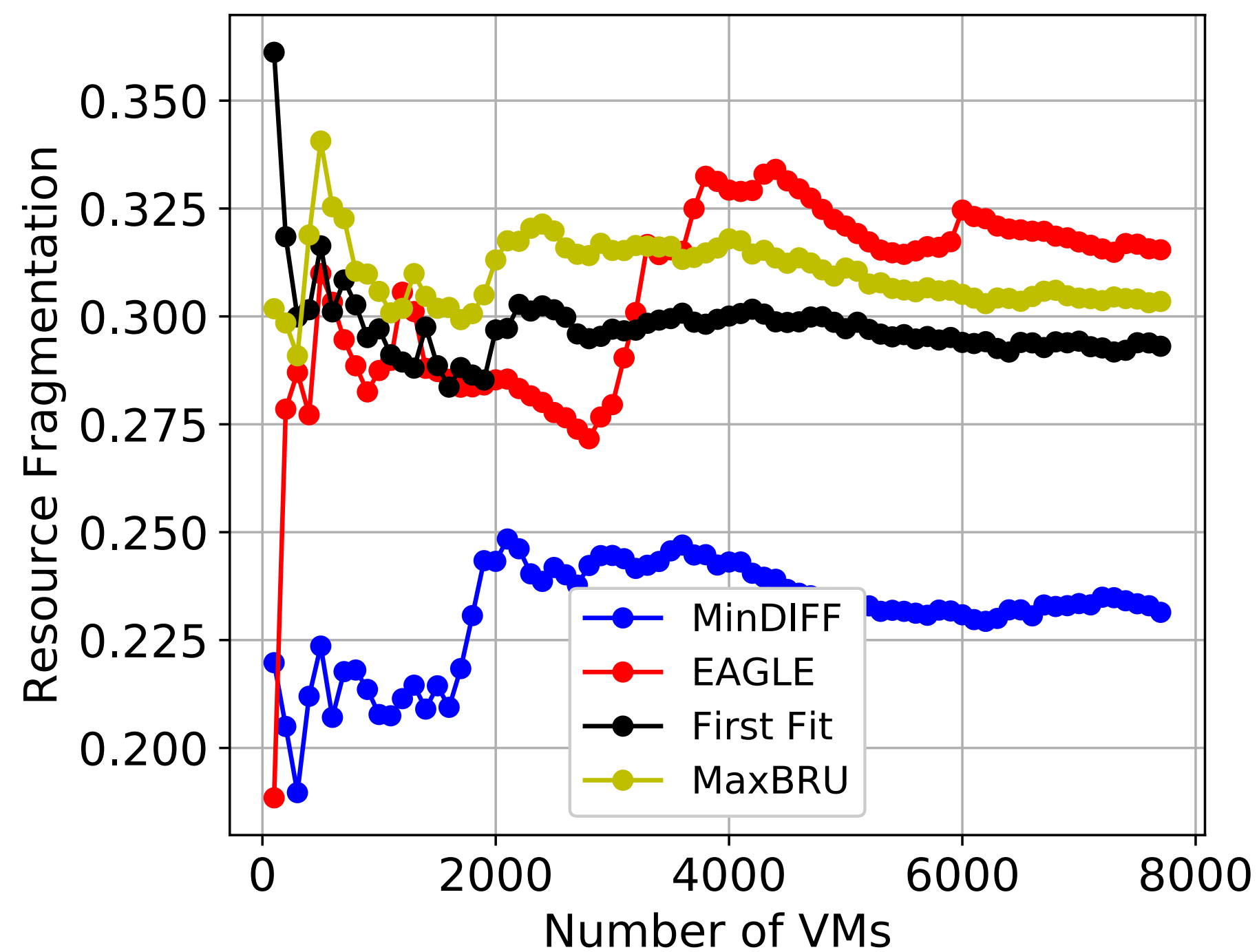
# Performance Evaluation

- ▶ If we do not consider the over commitment issue:
  - ▶ The number of used PMs and resource fragmentation



# Performance Evaluation

- ▶ The warning line is 80% along each dimension.
- ▶ Resource fragmentation and the percentage of PMs that CPU utilization is higher than 80%



Thank you!  
Q&A