LB-MAP: Load-Balanced Middlebox Assignment in Policy-Driven Data Centers

Alexander Ing, Manar Alqarni, Bin Tang, Mohsen Beheshti
aing1, malqarni1@toromail.csudh.edu, btang, mbeheshti@csudh.edu
Computer Science Department, California State University Dominguez Hills

Abstract

Middleboxes, such as firewalls and load balancers, are playing an increasingly important role in cloud data centers for security or performance purposes. In policy-driven cloud data centers, it requires that virtual machine (VM) traffic traverses a sequence of specified middleboxes in order to achieve security and performance guarantee. Much research has been done to study how to place virtualized middleboxes inside data centers for cost efficient VM traffic traversal. However, not much research has focused on load balance of middleboxes. In our research we study the Load-Balanced Middlebox Assignment Problem (LB-MAP), which minimizes the communication energy cost of VM pairs while satisfying their policy requirement as well as the capacity constraint of the switches that the middleboxes placed upon. We show that LB-MAP is equivalent to the classic minimum cost flow problem, which can be solved optimally and efficiently. We also design a suite of efficient heuristic algorithms based on different criteria VM-Based, MB-Based, and VM+MB-Based. Via extensive simulations, we show that all the heuristic algorithms perform close to the optimal minimum cost flow algorithm, while VM+MB-Based performs best among all the heuristic algorithms.

Introduction

A middlebox is a networking device that transforms, inspects, filters, or otherwise manipulates traffic for purposes other than packet forwarding. Traditional middlebox hardware is widely deployed in enterprise networks to improve network security and performance. The widespread deployment of middleboxes has resulted in some challenges and criticism due to poor performance and optimizations. To address above challenges, the integration of Software Defined Network (SDN) and Network Function Virtualization (NFV) has been recently proposed that enables efficient placement of software-based middlebox in commercial off the-shelf switches. Another problem facing datacenters is the communication cost within the datacenter. We measure the communication cost of any VM pair by focusing on the number of switches the communication goes through.

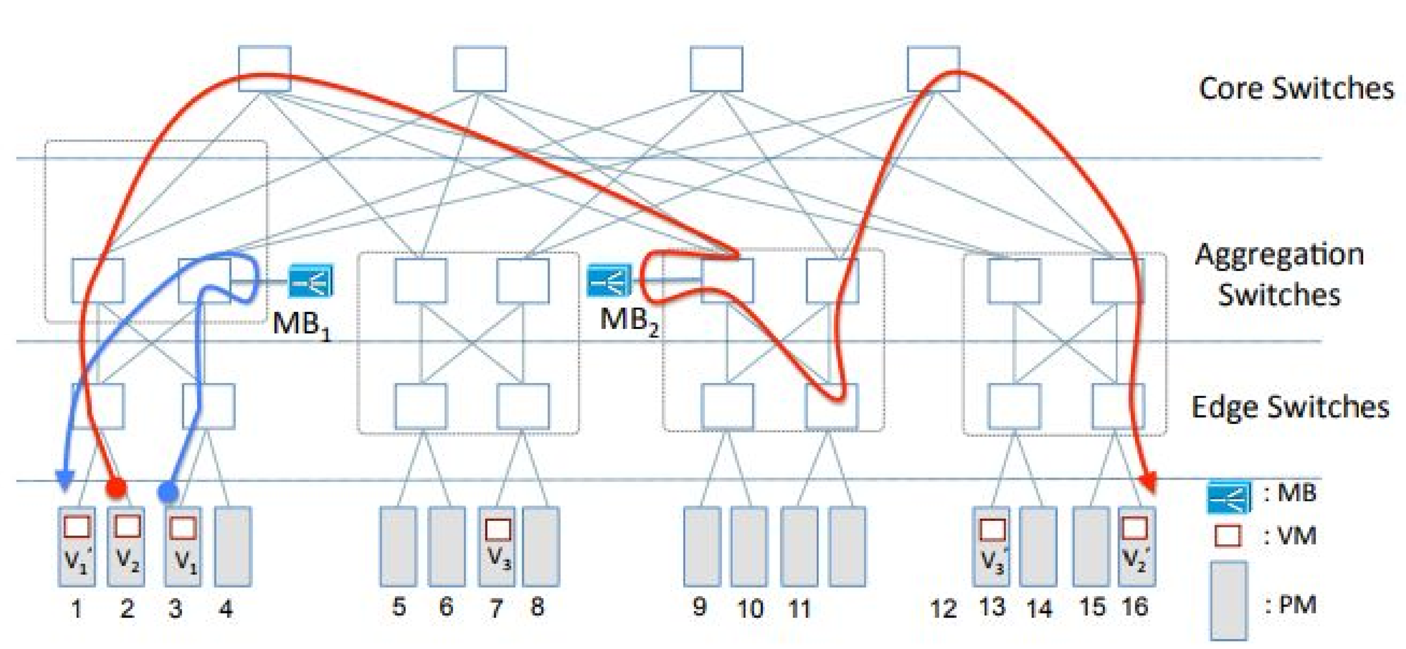


Figure 1

Problem Statement

- Given a datacenter with V communication pairs placed in the PMs
- Given M middleboxes with C capacity placed within the datacenter
- How to assign the flows from the communication pairs to the middleboxes such that:

➢ All traffic is assigned a middlebox
➢ No middlebox is assigned more than its capacity
➢ Total communication cost in the datacenter is minimized

Algorithms

VM-Based Algorithm
- Each VM pair will calculate cost of traversing all MB instances
- Assigned to MB that gives least cost
- Cannot assign a MB that reached its capacity

MB-Based Algorithm
- Each MB will calculate the cost of VM pairs
- Assign VM pairs until it reaches capacity
- Cannot assign VMs if it has reached capacity

VM-MB Based Algorithm
- Calculate cost of all VMs to every MB instance
- VMs assigned to MBs that result in least cost among all VM pairs, do so for all VM pairs
- MBs cannot exceed capacity

Minimum Cost Flow
- Classic optimization problem that finds the cheapest possible way of sending a certain amount of flow through a flow network
- Each edge between VM and MB contains cost

Simulation Parameters and Analysis

- *k = 4, 8, 16*
- *Number of servers = 16, 128, 1024*
- *p = 100, 200, 300, 400, and 500*
- *m = 1, 3, 5*
- VM+MB is closest to optimal minimum cost flow algorithm in all sets of parameters
- As communication pairs and middleboxes increase, the better MCF performs compared to the heuristics.

Figure 2

Minimum Cost Flow

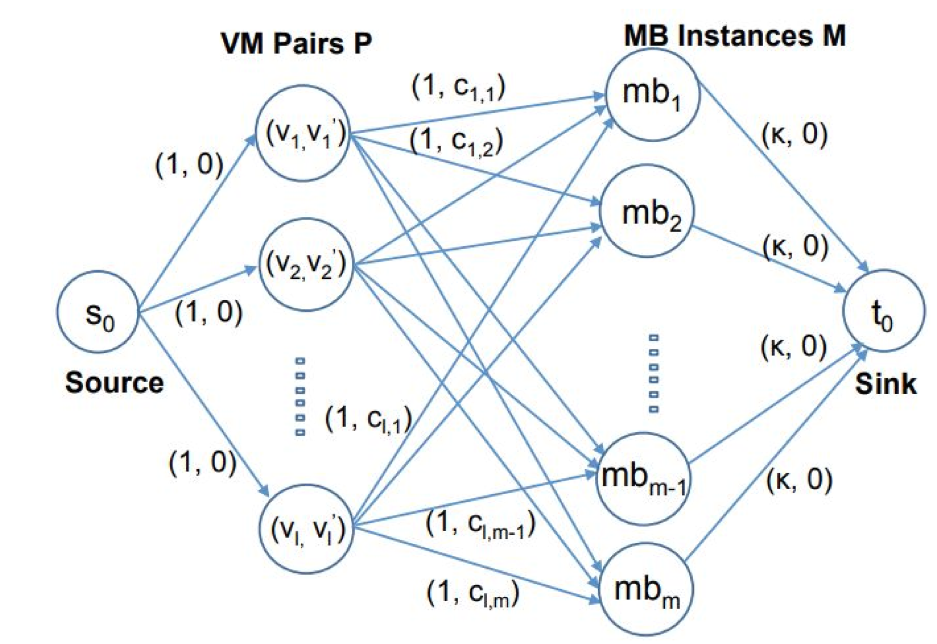


Figure 3

Varying number of MB instances under skewed energy model. VM pairs l = 300, number of PMs in data center is 128.

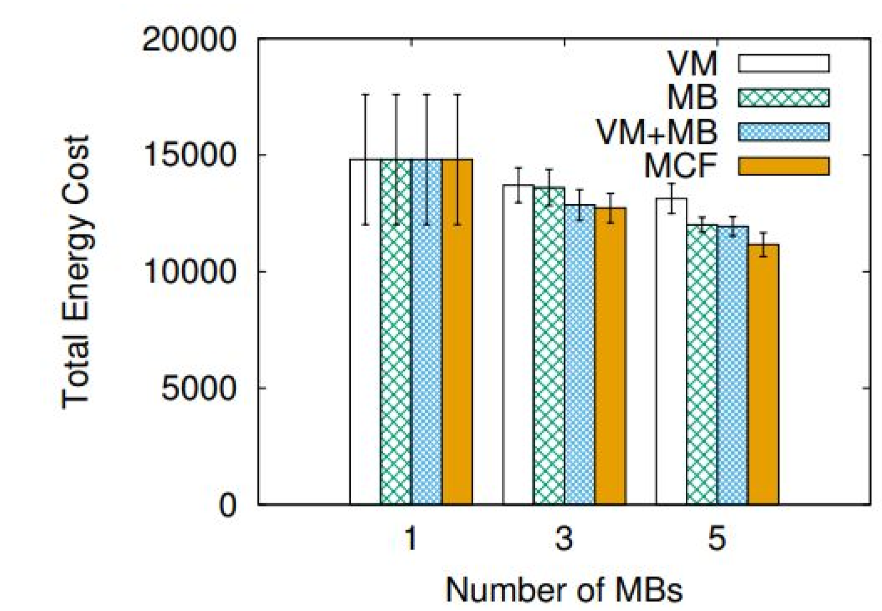
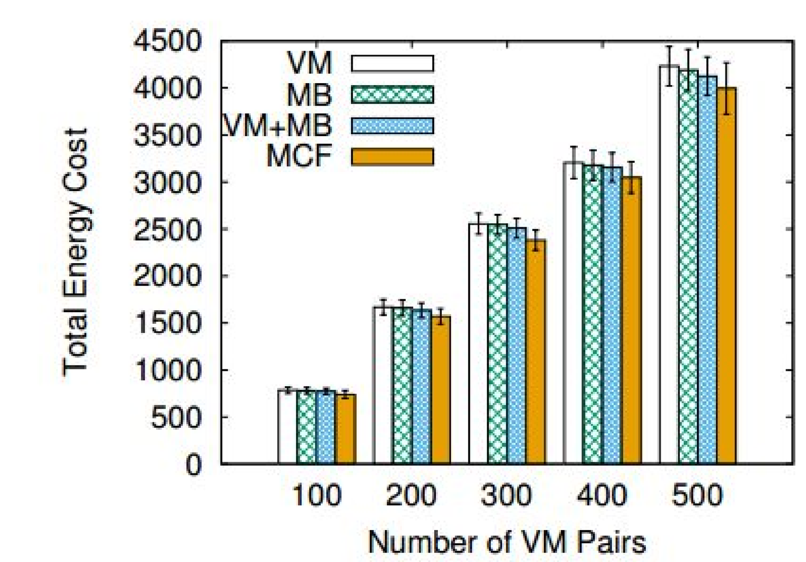


Figure 4

Varying number of VM pairs under uniform energy model. MBs m = 3, number of PMs in data center is 1024.



Conclusions

❑ We formulate energy-efficient middlebox placement problem and designed two time-efficient algorithms, in SDN-enabled data centers

❑ Extensive simulations show that VM+MB outperforms the other heuristic algorithms in all different network scenarios. Optimal minimum cost flow algorithm outperforms the heuristics.

❑ Future work: algorithm to support non-uniform middlebox load balancing.

Acknowledgements

This research is funded in part through the National Science Foundation (NSF) under Grant No.1649271 and CNS-1551221.