Green Data Center : Energy Saving

Reporter ：吳秉學
Advisor ： Hsueh-Wen Tseng

Paper introduction

- [1]HERO: Hierarchical energy optimization for data center networks
  - Yan Zhang; Ansari, N.
  - Communications (ICC), 2012 IEEE International Conference on , June 2012
- [2]Energy optimizations for data center network: Formulation and its solution
  - Shuo Fang; Hui Li; Chuan Heng Foh; Yonggang Wen; Khin Mi Mi Aung
  - Global Communications Conference (GLOBECOM), 2012 IEEE , Dec. 2012
- [3]GreenDCN: A General Framework for Achieving Energy Efficiency in Data Center Networks
  - Lin Wang; Fa Zhang; Arjona Aroca, J.; Vasilakos, A.V.; Kai Zheng; Chenying Hou; Dan Li; Zhiyong Liu
  - Selected Areas in Communications, IEEE Journal on ,January 2014

Outline

- Introduction
- Methods
  - Method 1
  - Method 2
  - Method 3
- Conclusions

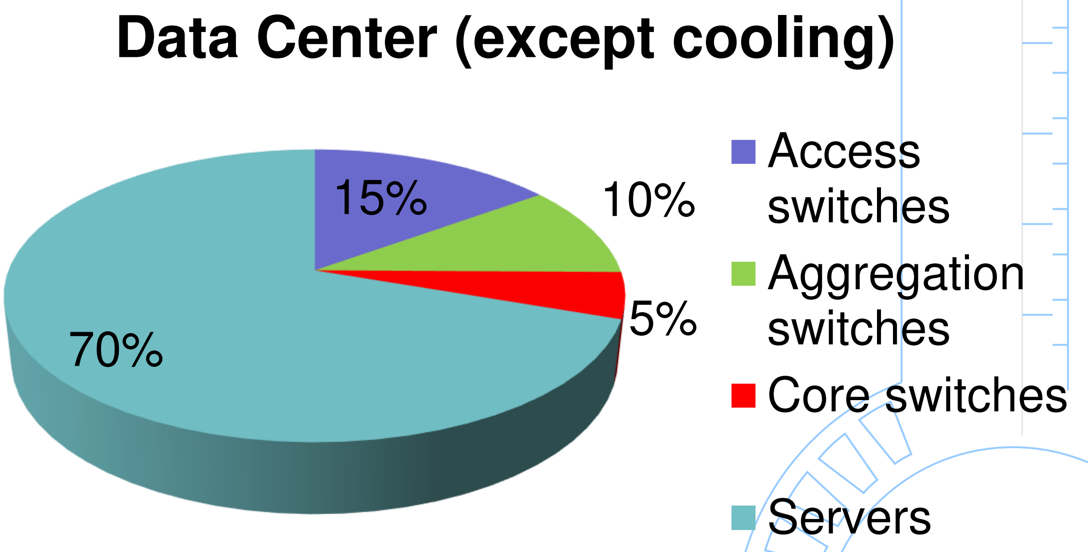
Introduction

- The energy consumption of data centers has become an essential problem
  - In 2013 U.S. data centers consumed an estimated 91 billion kwh of electricity [4]
  - Increase to roughly 140 billion kwh annually by 2020 [4]
    - $13 billion annually in electricity bills and 100 million metric tons of carbon pollution per year

[4]http://www.nrdc.org/energy/data-center-efficiency-assessment.asp

Introduction

- The main sources of power consumption in a data center
  - Cooling
  - Computing resources
  - Network elements



Cavdar, D.; Alagoz, F., "A survey of research on greening data centers," *Global Communications Conference (GLOBECOM), 2012 IEEE,* 2012

HERO: Hierarchical energy optimization for data center networks

Yan Zhang; Ansari, N.

Communications (ICC), 2012 IEEE International Conference on , June 2012

Method 1

- Data center networks become larger and larger
  - The complexity of solving this optimization problem increases
- **H**ierarchical **e**ne**r**gy **o**ptimization (HERO) model
  - Turning off some elements
  - Without violating the connectivity and QoS constraints

Method 1

- Five kinds of traffic
  - F1 : intra-edge switch traffic
  - F2 : inter-edge but intra-pod traffic
  - F3 : inter-pod traffic
  - F4 : incoming traffic
  - F5 : outgoing traffic

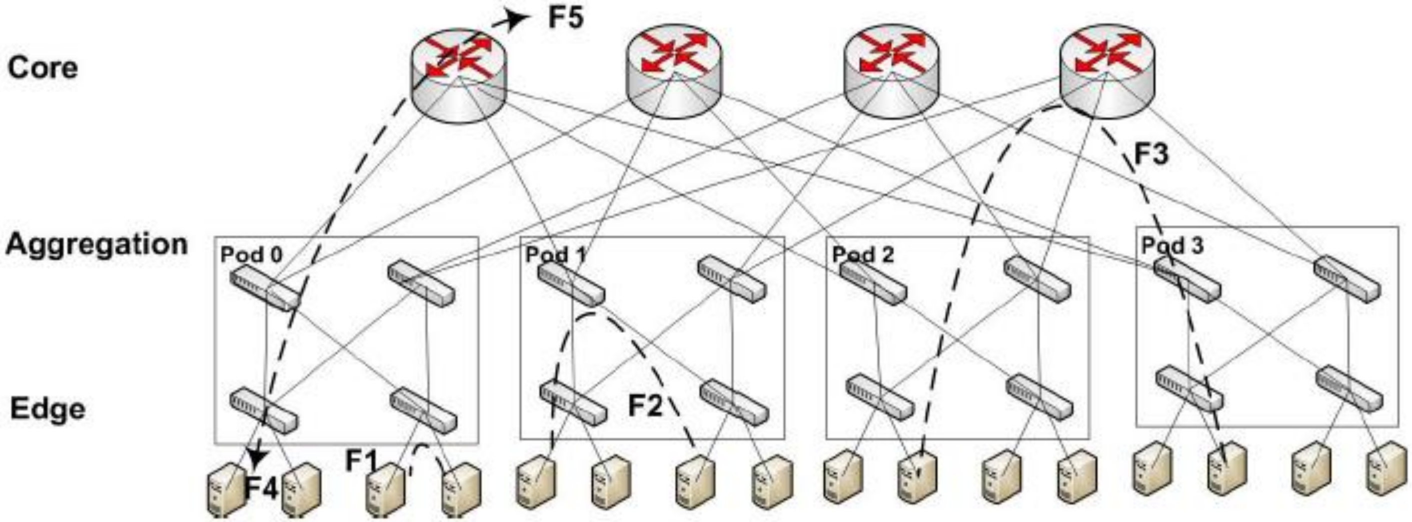


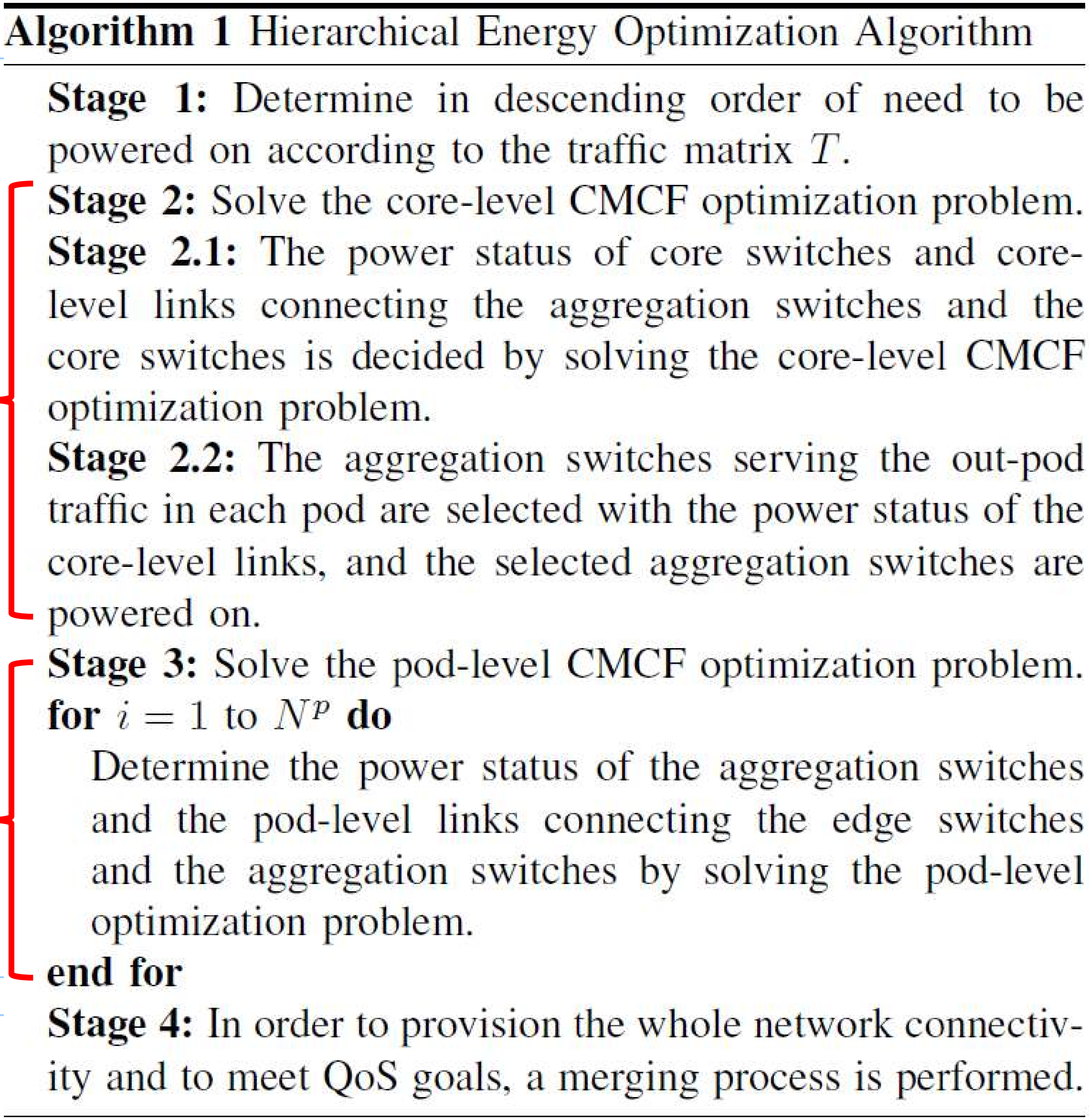
Fig. 1: Data center network topologies and traffic patterns

Method 1

- Two level power optimization
  - Core-level
    - Determine the core switches that must stay active to flow the outgoing traffic
    - Determine the aggregation switches which serve the out-pod traffic in each pod
  - Pod-level
    - Determine the aggregation switches that must be powered to flow the intra-pod traffic

CMCF : Capacitated Multi-commodity Minimum Cost Flow

Method 1



Algorithm 1 Hierarchical Energy Optimization Algorithm

```
Stage 1: Determine in descending order of need to be powered on according to the traffic matrix $T$.
Stage 2: Solve the core-level CMCF optimization problem.
Stage 2.1: The power status of core switches and core-level links connecting the aggregation switches and the core switches is decided by solving the core-level CMCF optimization problem.
Stage 2.2: The aggregation switches serving the out-pod traffic in each pod are selected with the power status of the core-level links, and the selected aggregation switches are powered on.
Stage 3: Solve the pod-level CMCF optimization problem.
for $i = 1$ to $N^p$ do
    Determine the power status of the aggregation switches and the pod-level links connecting the edge switches and the aggregation switches by solving the pod-level optimization problem.
end for
Stage 4: In order to provision the whole network connectivity and to meet QoS goals, a merging process is performed.
```

Core-level

Pod-level

Method 1

- Large traffic flows

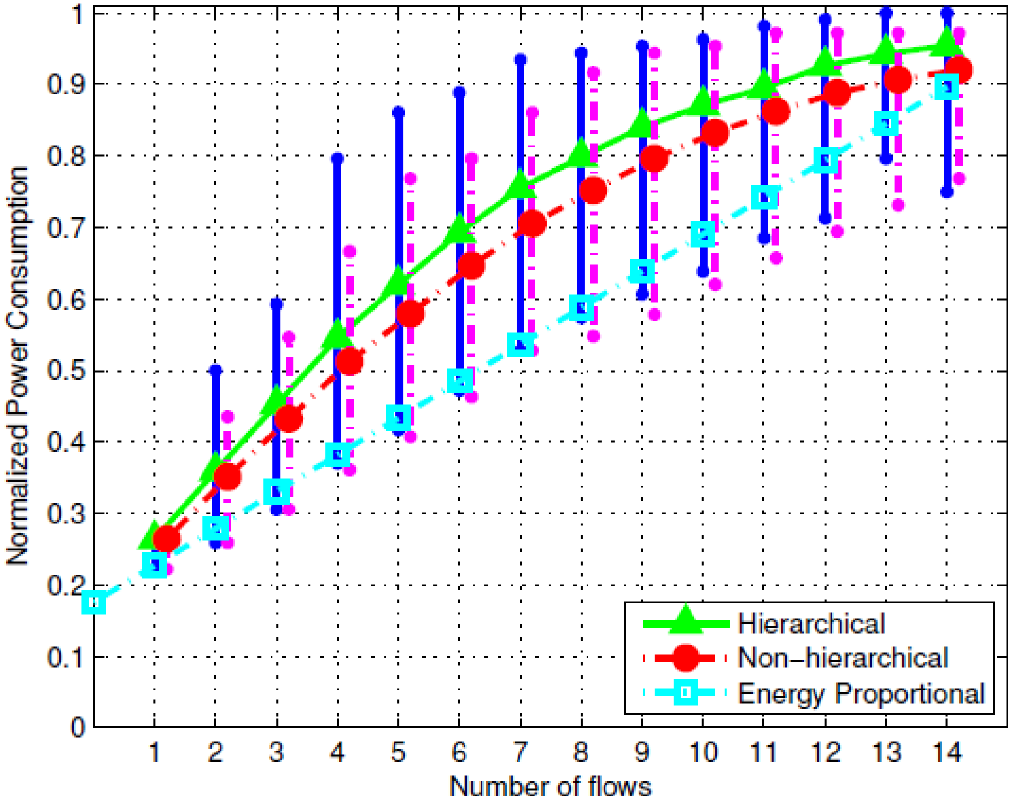


Fig. 3: The power consumption of 4-ary Fat-tree data center networks with different number of traffic flows.

Method 1

- Small traffic flows

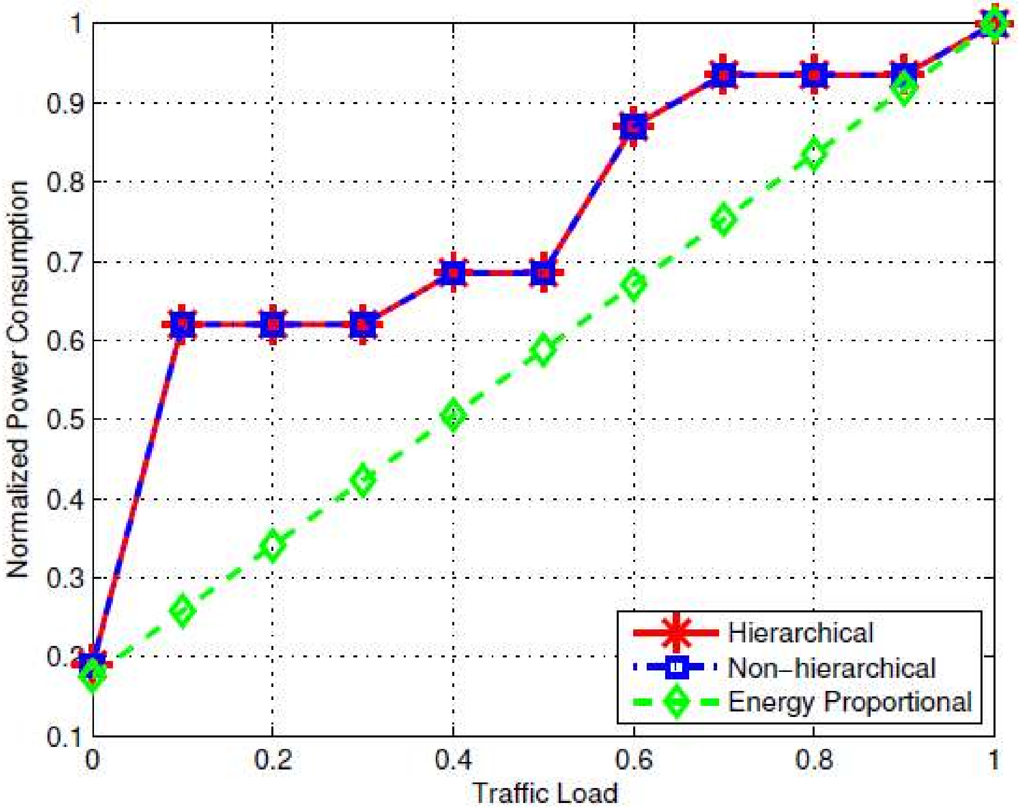


Fig. 4: The power consumption of a 4-ary Fat-tree data center network with all-to-all traffic under different traffic load.

Energy optimizations for data center network: Formulation and its solution

Shuo Fang; Hui Li; Chuan Heng Foh; Yonggang Wen; Khin Mi Mi Aung
Global Communications Conference (GLOBECOM), 2012 IEEE , Dec. 2012

Method 2

- Purpose
  - Minimize switch usage to save energy
  - Adjust link rates of switch ports according to traffic loads

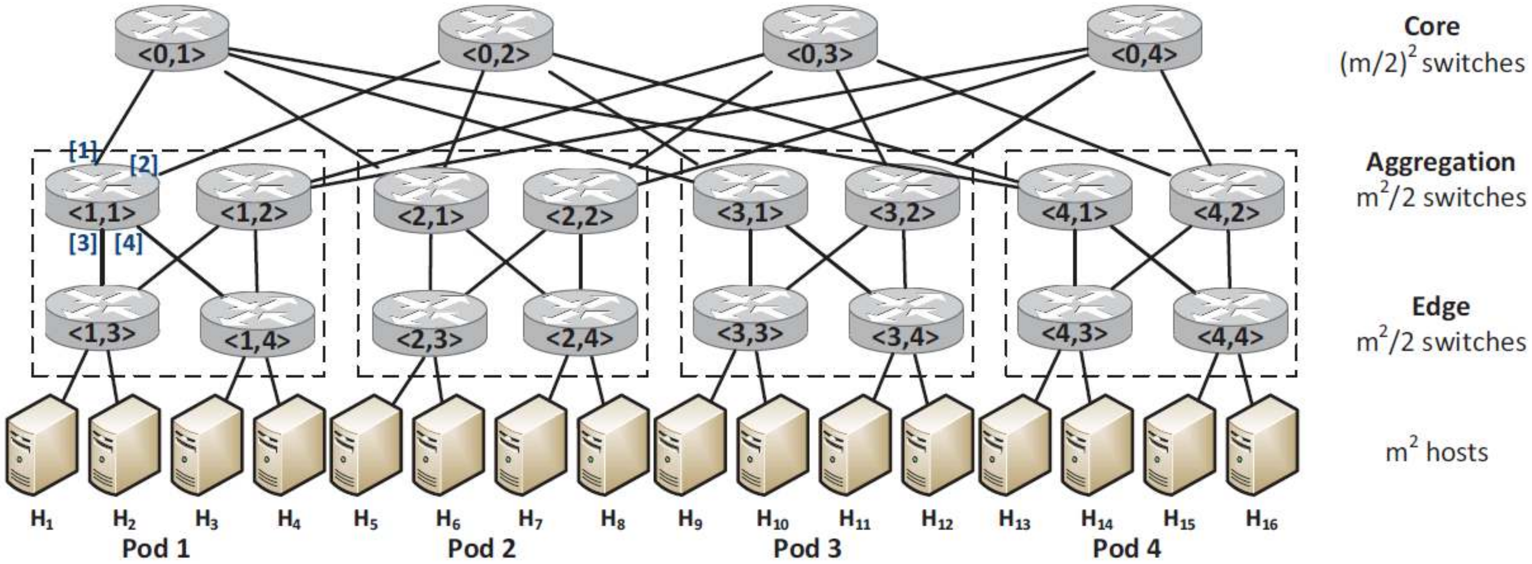


Fig. 1. Illustration of 4-ary Fat Tree topology.

Method 2

- Optimization formulation of the problem

Link rate

$$\min \sum_{\langle p,r\rangle} P\Big(\sum_{i} \boxed{l_i^{\langle p,r\rangle}}\Big)$$

subject to

Arrival load at incoming port k of switch <p,r>

$$\sum_{k} \boxed{\lambda_{k,d}^{\langle p,r\rangle}} = \sum_{k} \boxed{\mu_{k,d}^{\langle p,r\rangle}},$$

Traffic load at outgoing port k of switch <p,r>

$$\sum_{d} \lambda_{k,d}^{\langle p,r\rangle} \leq l_k^{\langle p,r\rangle}, l_i \in \mathcal{L},$$

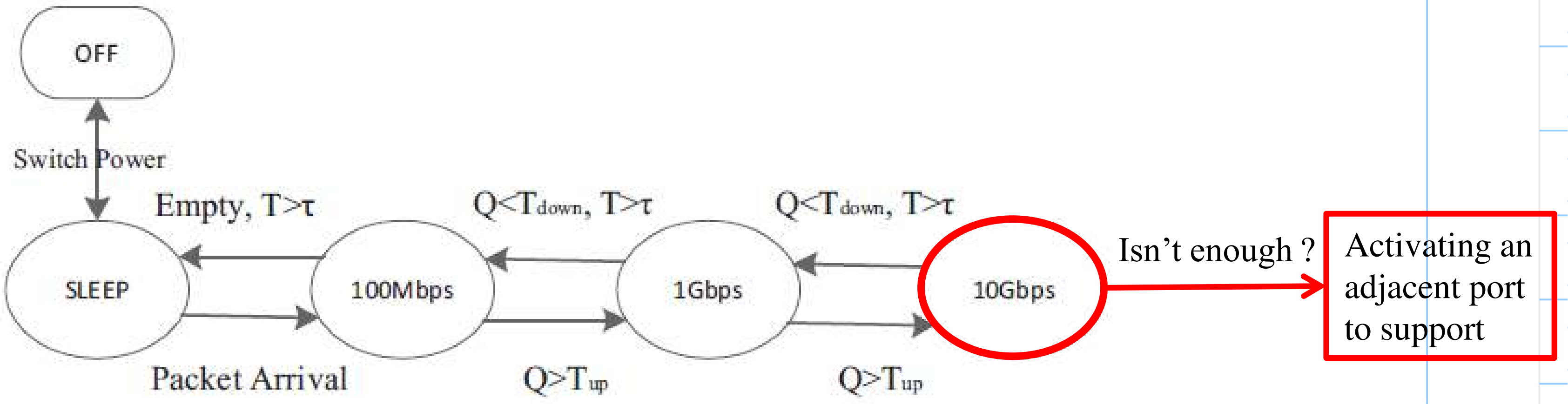
$$\sum_{d} \mu_{k,d}^{\langle p,r\rangle} \leq l_k^{\langle p,r\rangle}, l_i \in \mathcal{L},$$

Method 2

- Greedy approach
  - Utilize as few switches, switch links and switch link rates as possible
  - No active switches in the network system at the beginning
    - Switches are only enabled when packet arrives
  - Packets are automatically routed to a path on a spanning tree with the least link rate

Method 2

Port's view



Isn't enough ?

Activating an adjacent port to support

Fig. 4. Port state transition.

Q : Buffer level

T_{up} : upgrade threshold

T_{down} : downgrade threshold

τ : Time interval

Method 2

Switch's view

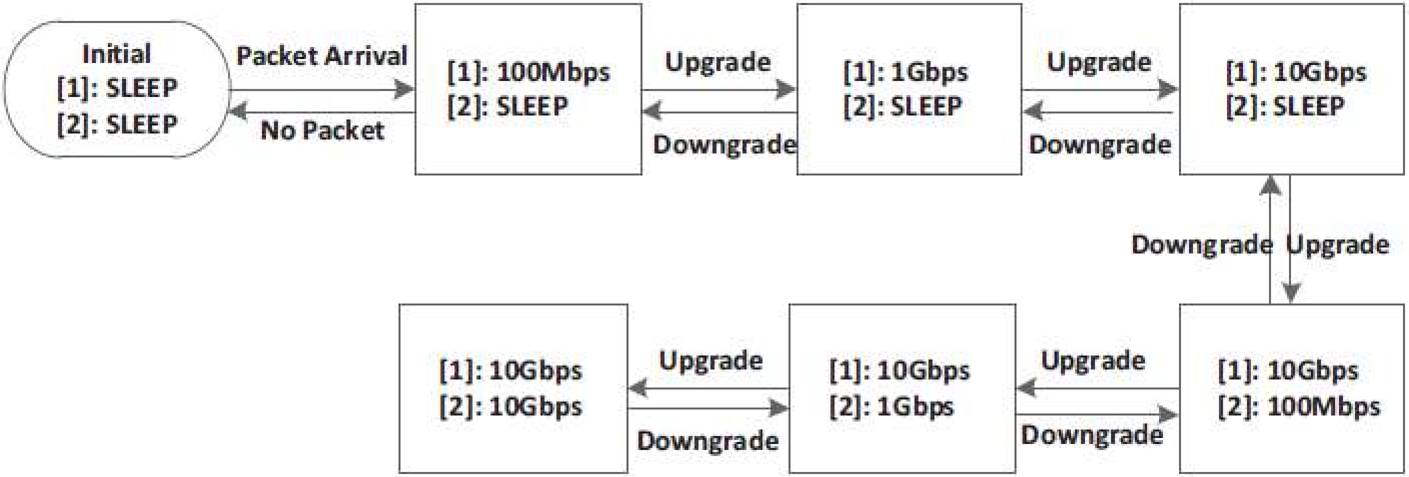


Fig. 5. Switch state transition.

Method 2

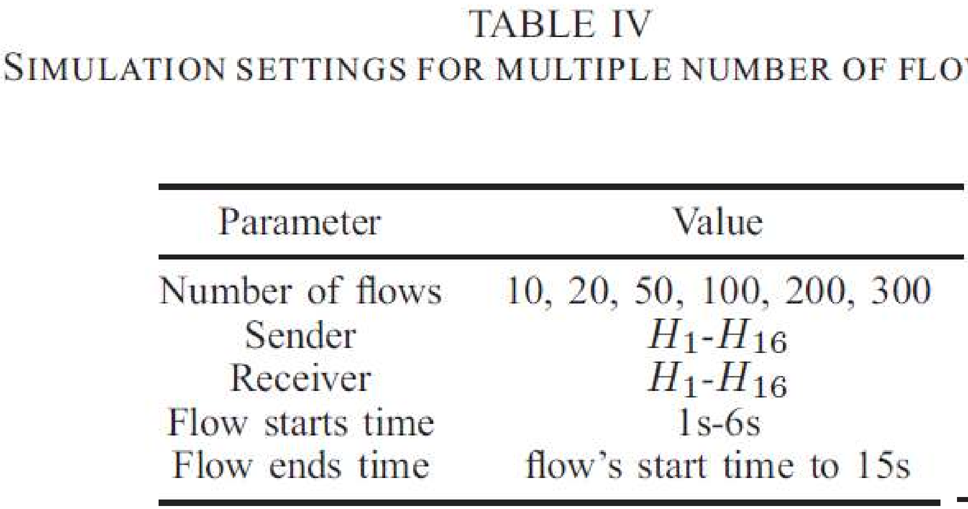


TABLE IV
SIMULATION SETTINGS FOR MULTIPLE NUMBER OF FLOWS TEST.

Parameter	Value
Number of flows	10, 20, 50, 100, 200, 300
Sender	H_1-H_{16}
Receiver	H_1-H_{16}
Flow starts time	1s-6s
Flow ends time	flow's start time to 15s

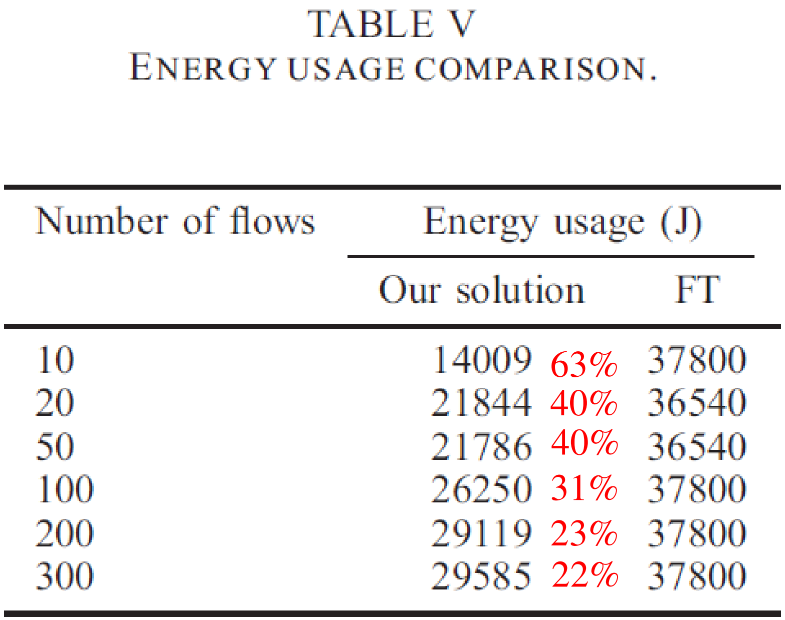


TABLE V
ENERGY USAGE COMPARISON.

Number of flows	Energy usage (J)		
	Our solution		FT
10	14009	63%	37800
20	21844	40%	36540
50	21786	40%	36540
100	26250	31%	37800
200	29119	23%	37800
300	29585	22%	37800

Method 2

Number of flows ↑, Delay↓

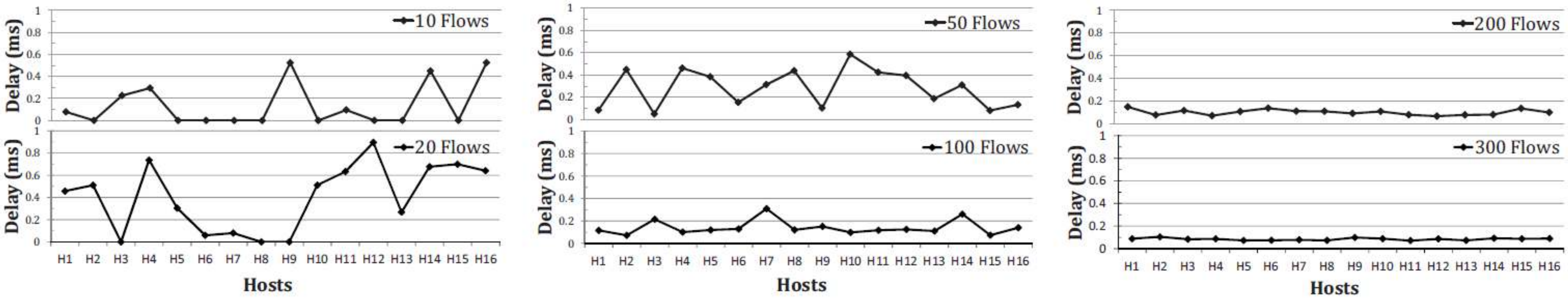


Fig. 9. Hosts delay statistics.

GreenDCN: A General Framework for Achieving Energy Efficiency in Data Center Networks

Lin Wang; Fa Zhang; Arjona Aroca, J.; Vasilakos, A.V.; Kai Zheng; Chenying Hou; Dan Li; Zhiyong Liu

Selected Areas in Communications, IEEE Journal on ,January 2014

Method 3

- In a typical data center from Google
  - The network power is approximately 20% of the total power when the servers are utilized at 100%
  - But it increases to 50% when the utilization of servers decreases to 15%

Method 3

- Purpose
  - Improve the energy efficiency in DCNs
- Explore unique features of data centers
  - Regularity of the topology
    - Fat-Tree, BCube and DCell
  - VM assignment
  - Application characteristics
- Design the VM assignment based on the applications' characteristics and regularity of the topology

Method 3

- General framework

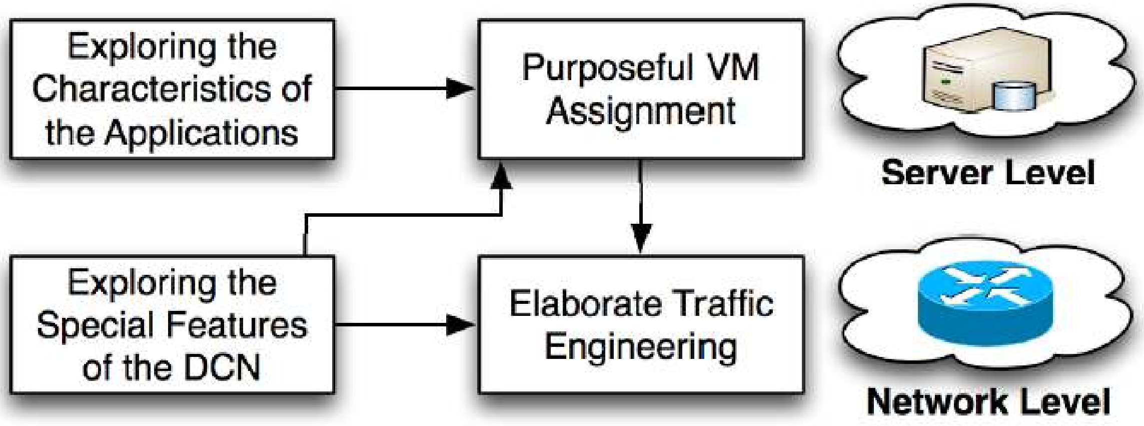


Fig. 1. A general framework for improving the energy efficiency in DCNs.

Method 3

- Modeling the energy-saving problem
  - Model by integer program

Energy curve for switch v

$$\min \ \sum_{v\in\mathcal{V}} \boxed{f(x_v)}$$

subject to

Total traffic going through node v, which never exceed the switch capacity C

$$\begin{aligned}
& x_v = \tfrac{1}{2}\sum_{e\in\mathcal{E}:e \text{ is incident to } v} y_e && \forall v \\
& x_v \le C && \forall v \\
& y_e = \sum_{d\in\mathcal{D}(t)} |d|\cdot\Phi_{d,e} && \forall e \\
& \Phi_{d,e} \in \{0,1\} && \forall d,e \\
& \Phi_{d,e}: \ \text{flow conservation} &&
\end{aligned}$$

Total load carried by link e

Whether the demand d goes through edge e

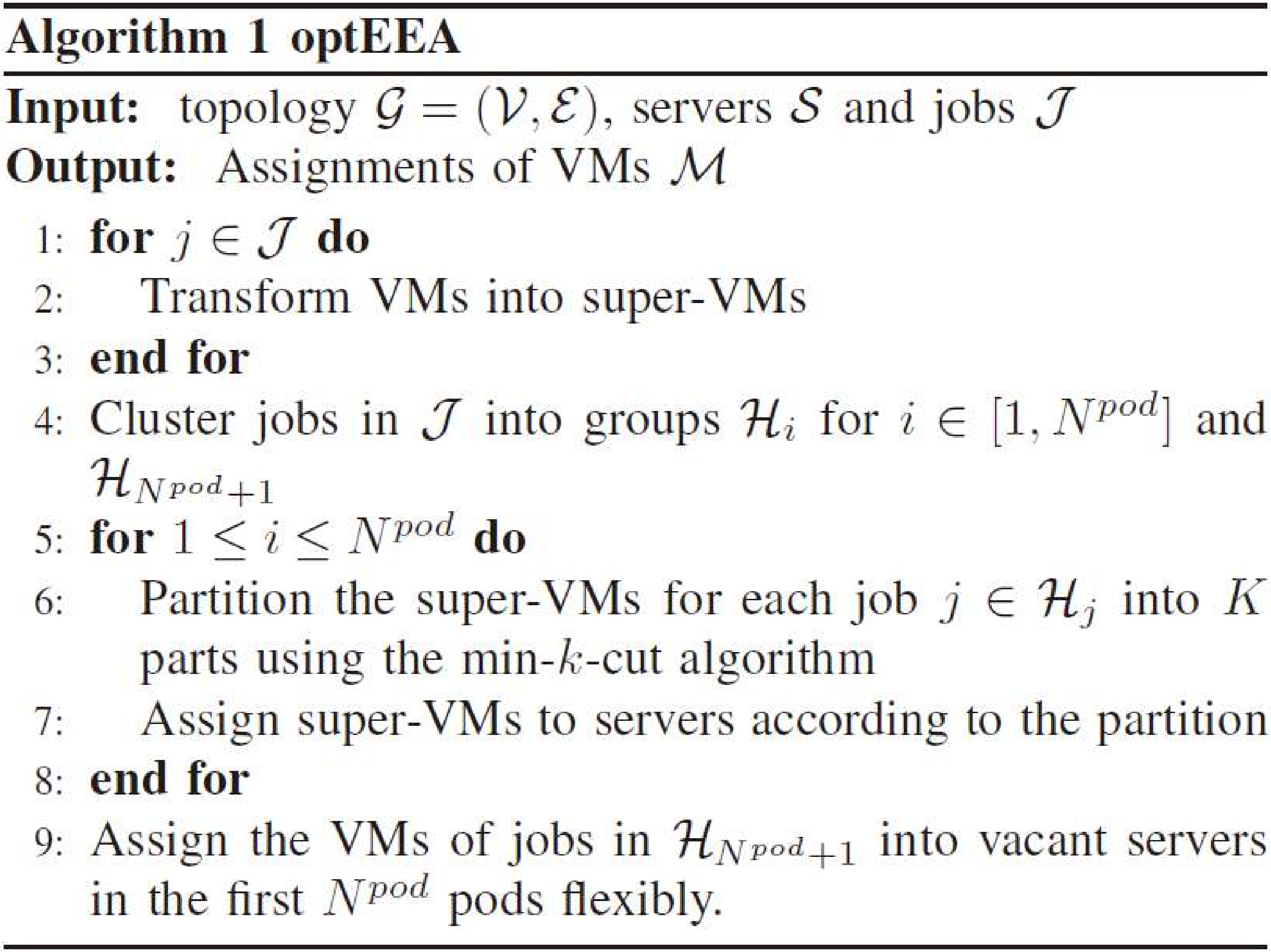
Theorem 1. Finding the optimality of the energy-saving problem in DCNs is NP-hard

Method 3

- Energy-efficient VM assignments
  - Three main principles for minimizing energy

1. At the rack level
   - Compacting VMs into racks as tightly as possible to minimize the power consumption of the ToR switches
2. At the aggregation level
   - Compacting VMs into a single rack is better than distributing the VMs into k racks
3. At the pod level
   - Same job, same pod

Method 3



Algorithm 1 optEEA

```
Input: topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, servers $\mathcal{S}$ and jobs $\mathcal{J}$
Output: Assignments of VMs $\mathcal{M}$
for $j \in \mathcal{J}$ do
    Transform VMs into super-VMs
end for
Cluster jobs in $\mathcal{J}$ into groups $\mathcal{H}_i$ for $i \in [1, N^{pod}]$ and $\mathcal{H}_{N^{pod}+1}$
for $1 \leq i \leq N^{pod}$ do
    Partition the super-VMs for each job $j \in \mathcal{H}_j$ into $K$ parts using the min-$k$-cut algorithm
    Assign super-VMs to servers according to the partition
end for
Assign the VMs of jobs in $\mathcal{H}_{N^{pod}+1}$ into vacant servers in the first $N^{pod}$ pods flexibly.
```

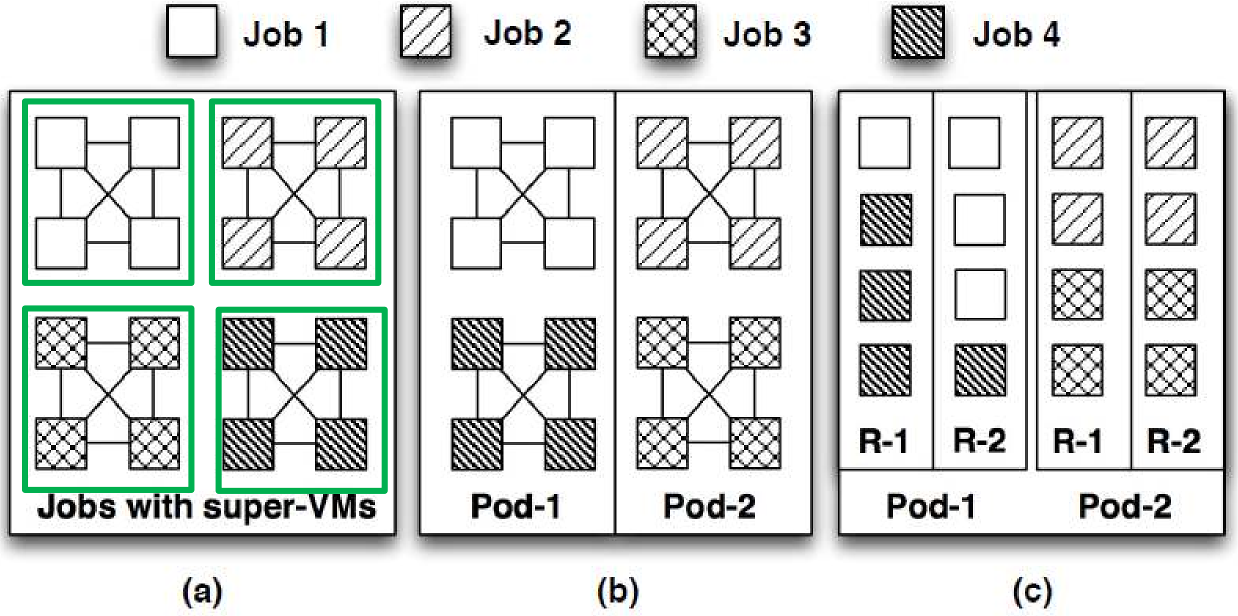
Step 1.

Step 2.

Step 3.

Method 3

- Algorithm 1- example

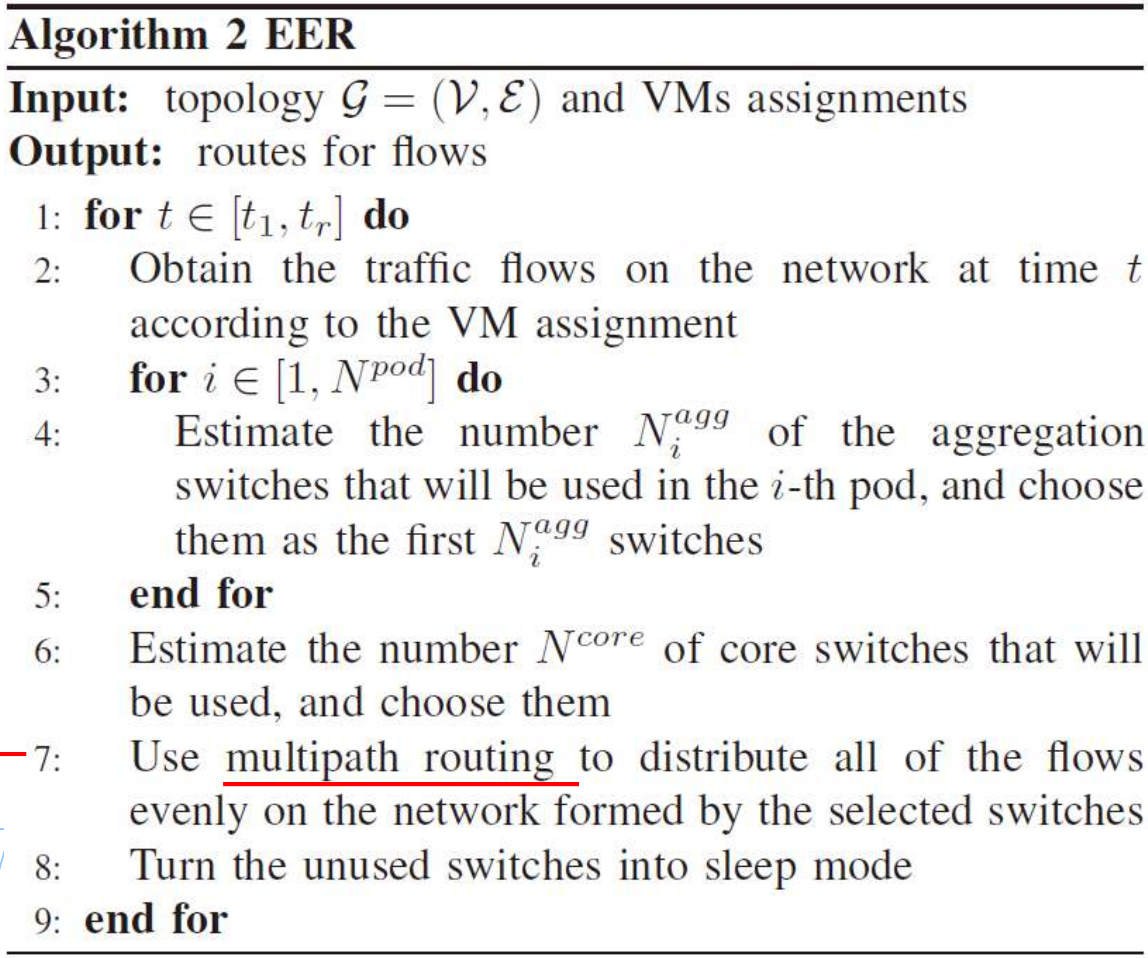


Super-VM

Fig. 2. (a) Original jobs' VMs are transformed to super-VMs; (b) the resulting super-VMs are clustered into pods using the k-means clustering algorithm; (c) after assigning jobs to pods, the super-VMs are assigned to racks using the minimum k-cut algorithm.

Method 3

- Energy-efficient Routing



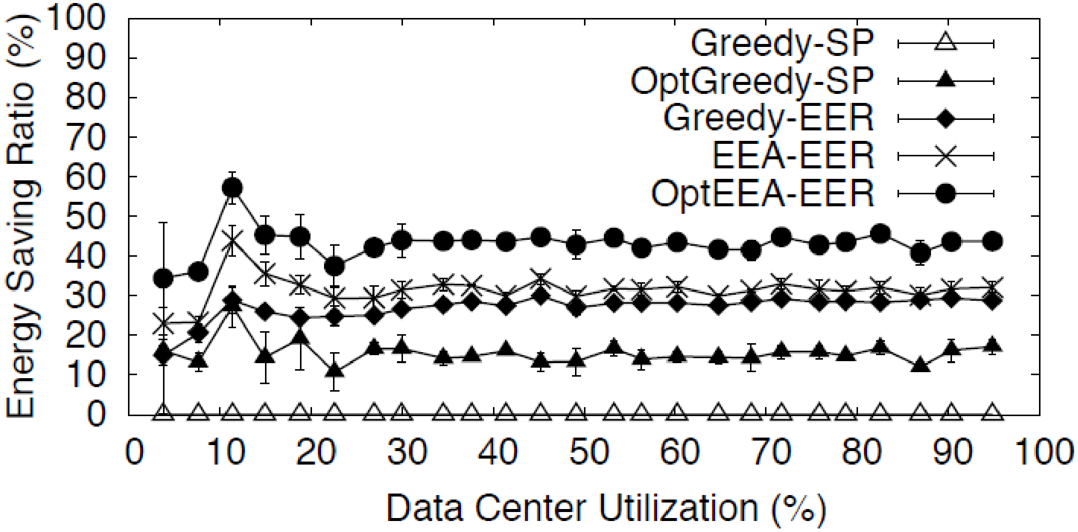
Algorithm 2 EER

```
Input: topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and VMs assignments
Output: routes for flows
for $t \in [t_1, t_r]$ do
    Obtain the traffic flows on the network at time $t$ according to the VM assignment
    for $i \in [1, N^{pod}]$ do
        Estimate the number $N_i^{agg}$ of the aggregation switches that will be used in the $i$-th pod, and choose them as the first $N_i^{agg}$ switches
    end for
    Estimate the number $N^{core}$ of core switches that will be used, and choose them
    Use multipath routing to distribute all of the flows evenly on the network formed by the selected switches
    Turn the unused switches into sleep mode
end for
```

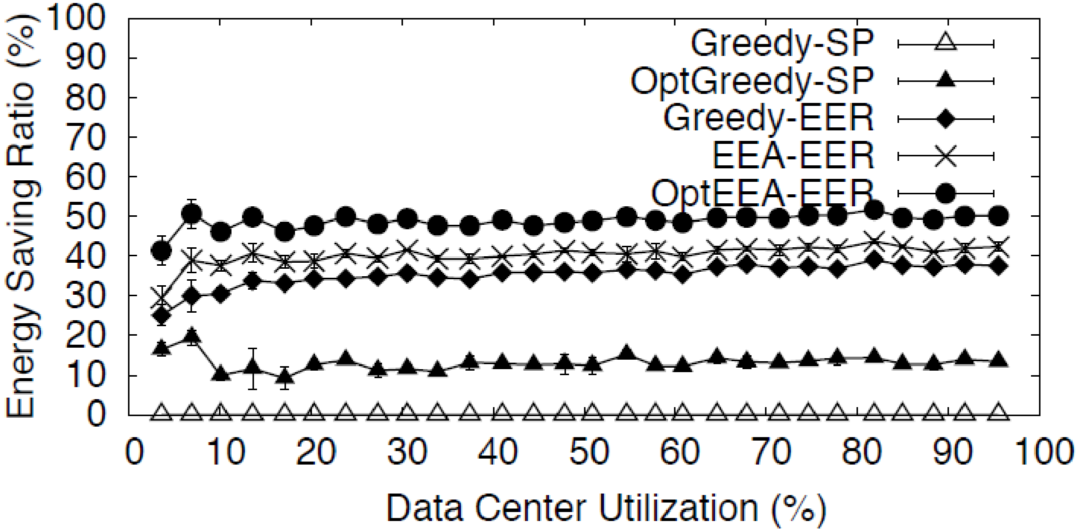
MPTCP

Method 3

Efficiency of Energy Saving



(a) 320 switches

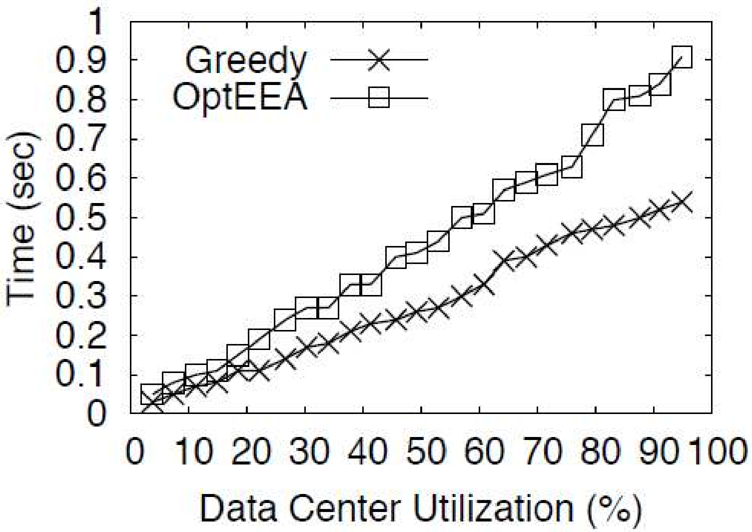


(b) 720 switches

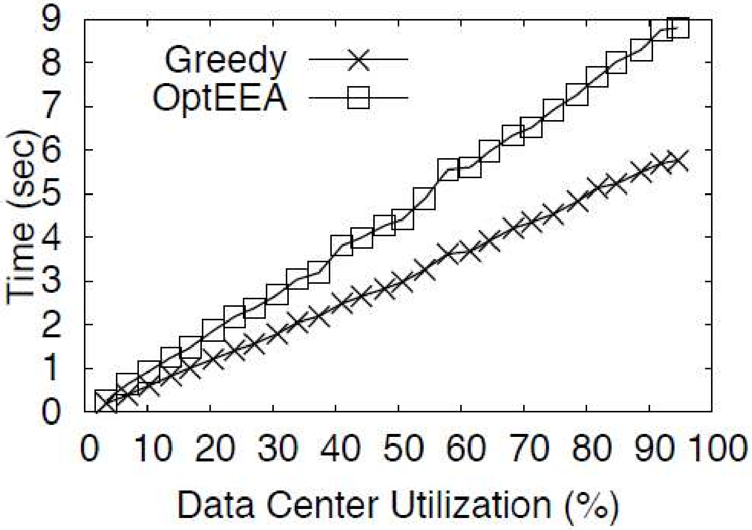
(Normalized by the Greedy-SP result)

Method 3

Running Time



(a) 320 switches



(b) 720 switches

Conclusions

- Turn on/off the switches [1][2][3]
- VM assignment and energy-efficient routing[3]
- Power saving
  - IT equipment
    - Server [3]
    - Switch [1][2][3]

Conclusions

	Resource allocation	Power allocation	Topology	QoS	Load Balance
[1]		○	Fat-Tree	○	○
[2]	○	○	Fat-Tree		○
[3]	○	○	Fat-Tree		○