Cloud computing: state-of-the-art and research challenges

Q. Zhang, L. Cheng, and R. Boutaba, J. Internet Services and Applications, 2010

DENS: Data Center Energy-Efficient Network-Aware Scheduling

D. Kliazovich, P. Bouvry, and S. U. Khan, IEEE/ACM Int Conf on Green Computing and Communications, 2010

Improving the Scalability of Data Center Networks with Traffic-aware Virtual Machine Placement

X. Meng, V. Pappas, and L. Zhang, In Proc of INFOCOM, 2010

100-2

Data Center Network

2012/05/29

蔡欣哲 王郁筑

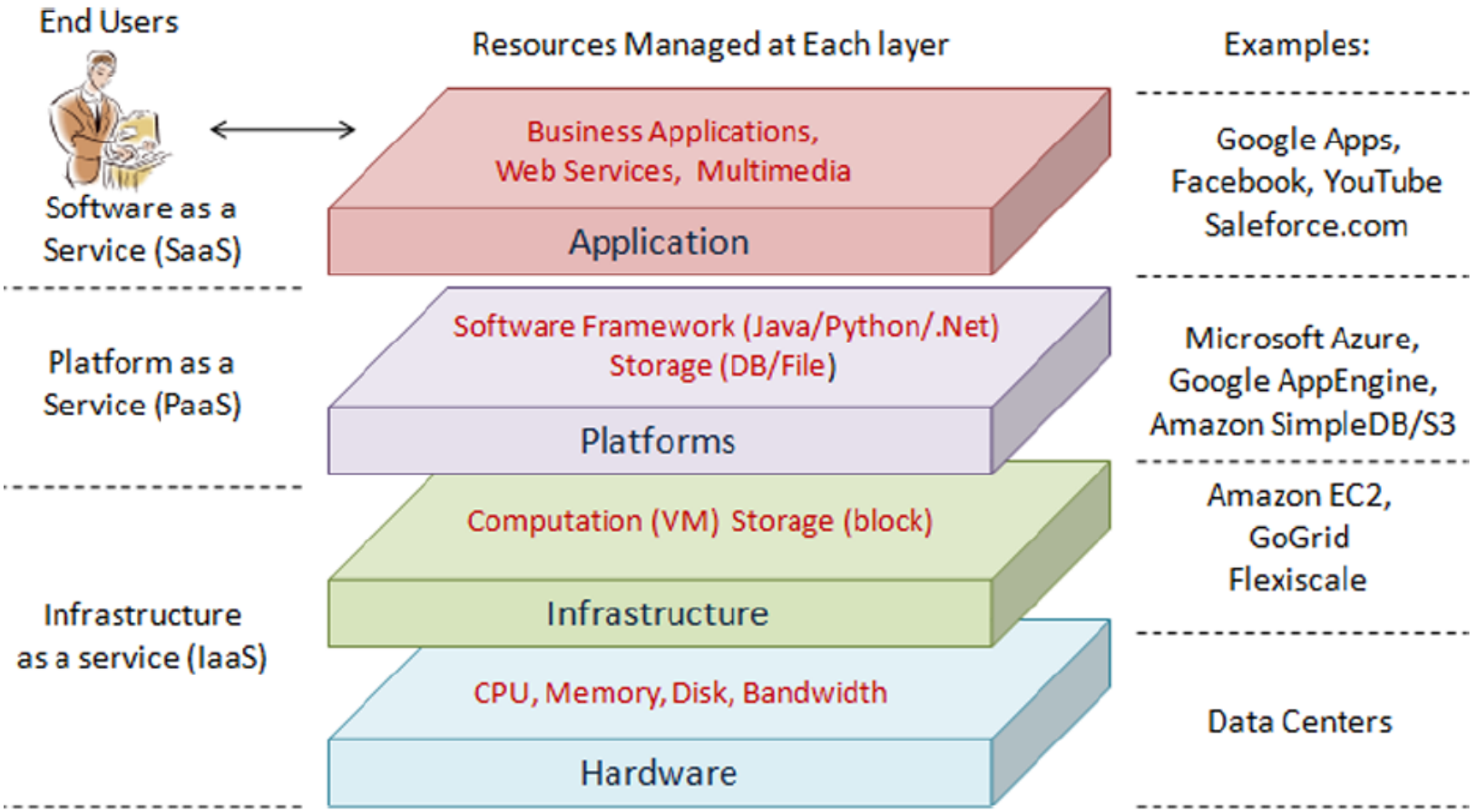
Paper Study-1

1) “Cloud computing: state-of-the-art and research challenges”

- Introduction
- Research Challenges
  - summarize the current research topics in cloud computing

Introduction(1/3)

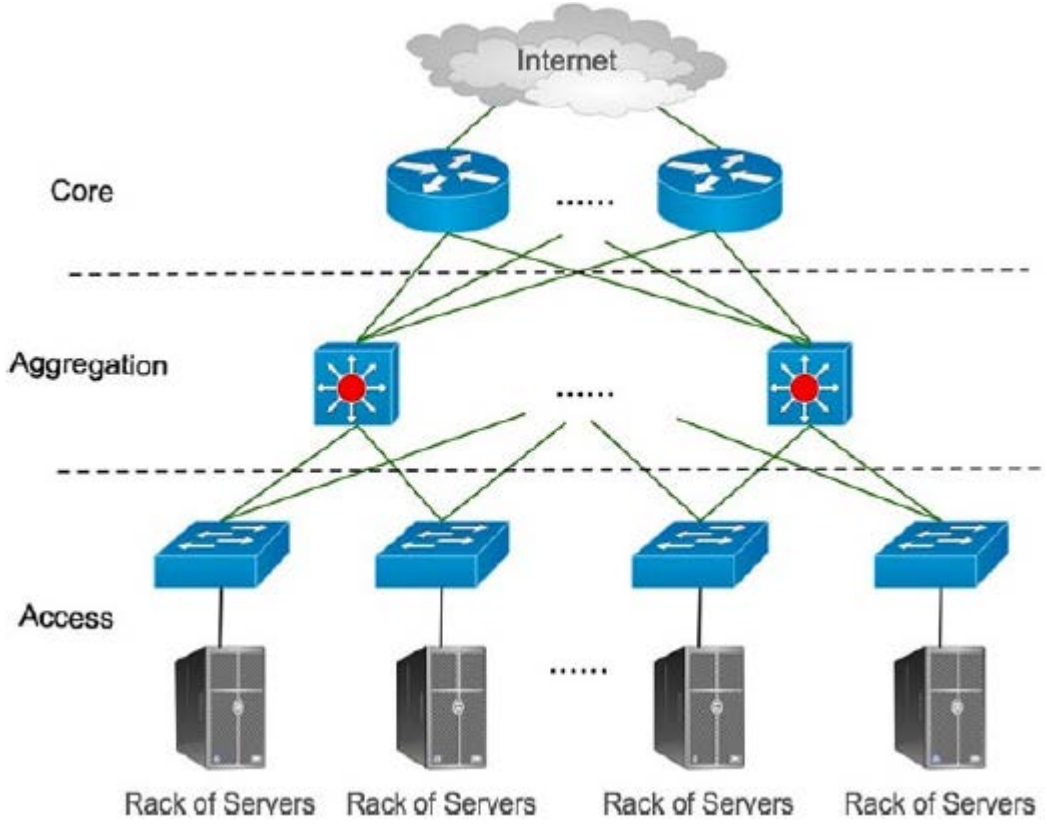
- Cloud computing environment can be divided into 4 layers.
  - hardware/datacenter, infrastructure, platform, application



Cloud computing architecture

Introduction(2/3)

- Architectural design of data centers
  - A data center, which is home to the power and storage, is central to cloud computing and contains thousands of devices like servers, switches and routers.



Basic layered design of data center network infrastructure

Introduction(3/3)

- The design of data center network architecture should meet the following objectives.
  - Uniform high capacity
  - Free VM migration
  - Resiliency
  - Scalability
  - Backward compatibility
  - ……
- We summarize some of the challenging research issues in cloud computing (especially for data center).

Research challenges(1/5)

- Server consolidation
  - It is an effective approach to maximize resource utilization while minimizing energy consumption.
    - Ex: Live VM migration
  - Optimally consolidating servers in data center is often formulated as an NP-hard optimization problem.
  - Server consolidation activities should not hurt application performance.
    - Ex: Maximally consolidating a server may result in resource congestion when a VM changes its footprint on the server.

Research challenges(2/5)

- Energy management
  - Cost of powering and cooling accounts for 53% of the total operational expenditure of data centers.

- Designing energy-efficient data centers has recently received considerable attention.
  - Energy-efficient hardware architecture
  - Energy-aware job scheduling and server consolidation
  - Network protocols and infrastructures

- A key challenge is to achieve a good trade-off between energy savings and application performance.

Research challenges(3/5)

- Traffic management and analysis
  - Analysis of data traffic is important for today's data centers.
    - Ex: Web applications rely on analysis of traffic data to optimize customer experiences.
  - There are several challenges for existing methods.
    - Ex: Existing methods can compute traffic between a few hundreds end hosts, but even a modular data center can have several thousand servers.
  - Currently, there is not much work on measurement and analysis of data center traffic.

Research challenges(4/5)

- Data security
  - Service providers do not have access to physical security system of data centers
    - They rely on infrastructure provider to achieve full data security.
  - Infrastructure provider must achieve the following objectives.
    - Confidentiality: using cryptographic protocols.
    - Auditability: using remote attestation techniques.
  - It is critical to build trust mechanisms at every architectural layer.
    - Ex: hardware trusted platform module(TPM), secure virtual machine monitors

Research challenges(5/5)

- Novel cloud architectures
  - Most of the commercial clouds are implemented in large data centers.
    - (adv) economy-of-scale and high manageability
    - (disadv) high energy expense and high initial investment
  - Recent work suggests that small size data centers can be more advantageous in many cases.
    - does not consume so much power
    - cheaper to build and better geographically distributed
  - GreenCloud architecture

Paper Study-2

2) "DENS: Data Center Energy-Efficient Network-Aware Scheduling"

- Introduction
- Background
  - data center architecture, energy consumption models, and data center network congestion
- The "DENS" Methodology
  - presents the core of the scheduling approach
- Performance Evaluation
  - discuss experimental results
- Conclusions

Introduction(1/2)

- Energy consumption is a growing concern for data centers operators.

- Two popular techniques for power savings
  - Dynamic Voltage and Frequency Scaling(DVFS)
    - adjusts hardware power consumption according to applied computing load
  - Dynamic Power Management(DPM)
    - achieves most of energy savings by powering down devices at runtime

- To make DPM scheme efficient
  - a scheduler must consolidate data center jobs on a minimum set of computing resources to maximize the amount of unloaded servers that can be powered down(or put to sleep)

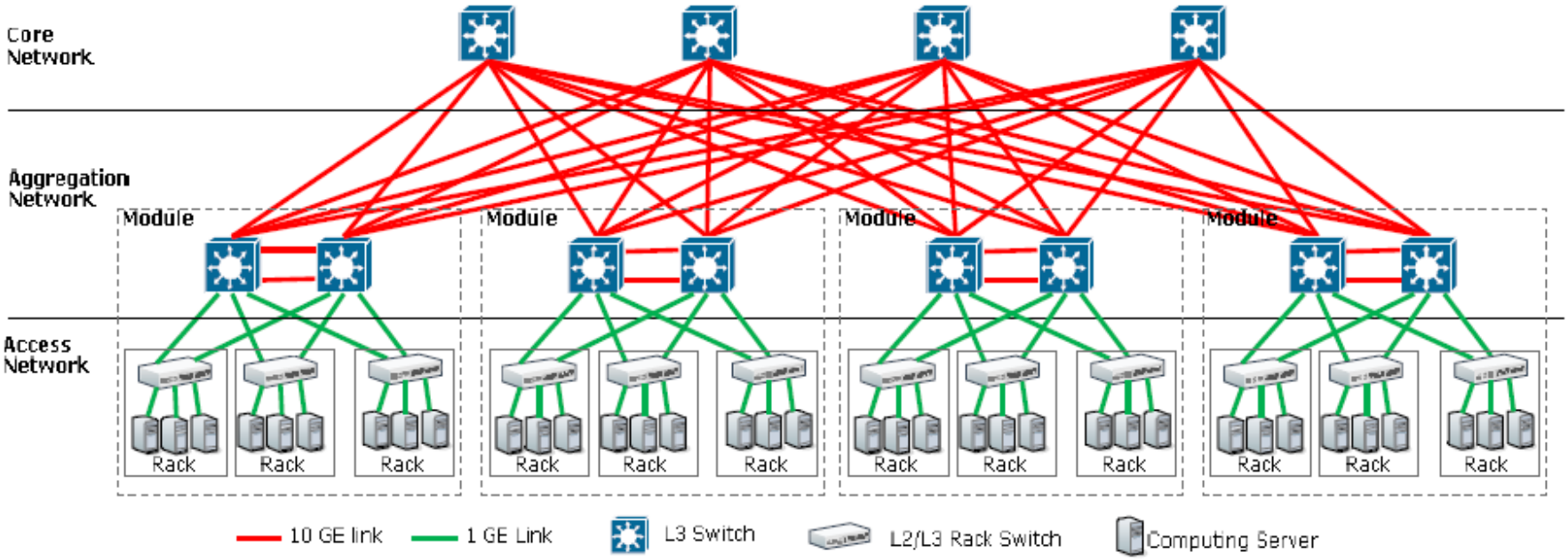
Introduction(2/2)

- This paper presents a scheduling methodology .
  - **D**ata center **E**nergy-efficient **N**etwork-aware **S**cheduling(DENS)

- DENS methodology
  - achieve the balance between individual job performances, traffic demands, and energy consumed by the data center
  - avoid hotspots within a data center while minimizing the number of computing servers required for job execution

Background(1/3)

- Data Center Topology
  - Three-tier trees of hosts and switches form the most widely used data center architecture.
    - Core tier, aggregation tier and access tier
    - Racks are arranged in modules with a pair of aggregation switches



Three-tier data center architecture

Background(2/3)

- Energy Models
  - Computing servers account for a major portion of data center energy consumption.
    - Two main approaches for reducing energy consumption in computing servers.(DVFS and DPM)
  - Switches that delivers job requests to the computing servers for execution.
    - Not all of switches can dynamically be put to sleep.(Ex: core switch)
    - Aggregation switches service modules, which can be powered down when module racks are inactive.

Background(3/3)

- Data Center Network Congestion
  - Small buffers and the mix of high-bandwidth traffic are the main reasons for network congestion.
  - Congestion(or hotspots) may affect the ability of a data center network to transport data.
    - Any of the switches may become congested either in uplink direction or downlink direction or both.
  - To benefit the three-tiered architecture, jobs must be distributed among the computing servers.
    - It contradicts the objectives of energy-efficient scheduling.
    - How to tradeoff ?

The DENS Methodology(1/3)

- DENS minimizes the total energy consumption of a data center.
  - By selecting best-fit computing resources for job execution based on the load level and communication potential.

- Contrary to traditional scheduling solutions
  - For a three-tier data center, DENS metric M is defined as a weighted combination of server-level f_s , rack-level f_r , and module-level f_m .

$$M = \alpha \cdot f_s + \beta \cdot f_r + \gamma \cdot f_m$$

where α, β and γ are weighted coefficients

The DENS Methodology(2/3)

- The factor related to the choice of computing servers combines server load $L_s(l)$ and its communication potential $Q_r(q)$.

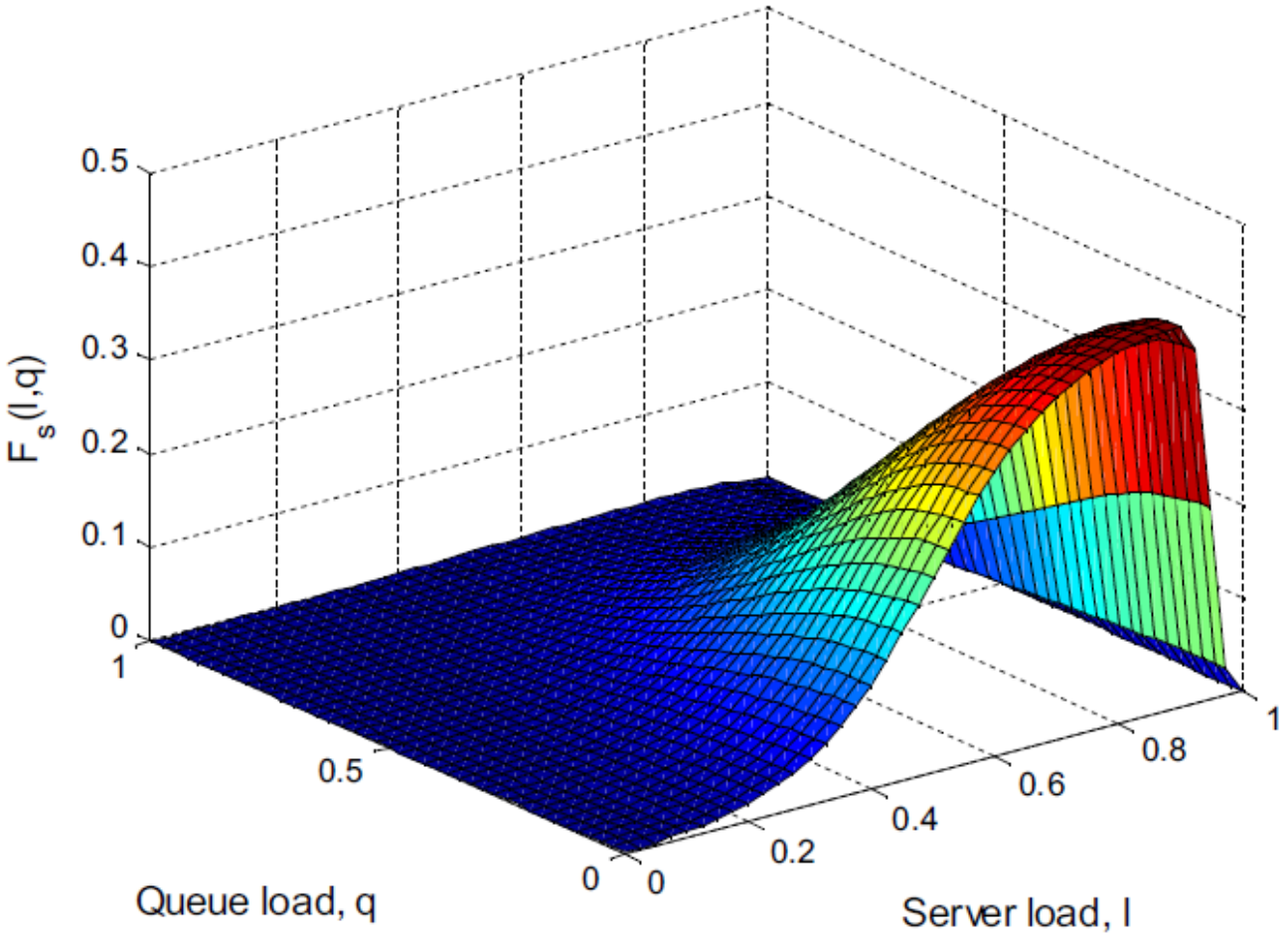
$$f_s(l,q) = L_s(l) \cdot \frac{Q_r(q)^{\varphi}}{\delta_r}$$

$L_s(l)$: *the load of the individual servers* l

$Q_r(q)$: *the load at the rack uplink by analyzing the congestion level in the switch's outgoing queue* q

The DENS Methodology(3/3)

- The obtained bell-shaped favors selection of servers with the load level above average located in racks with the minimum or no congestion.

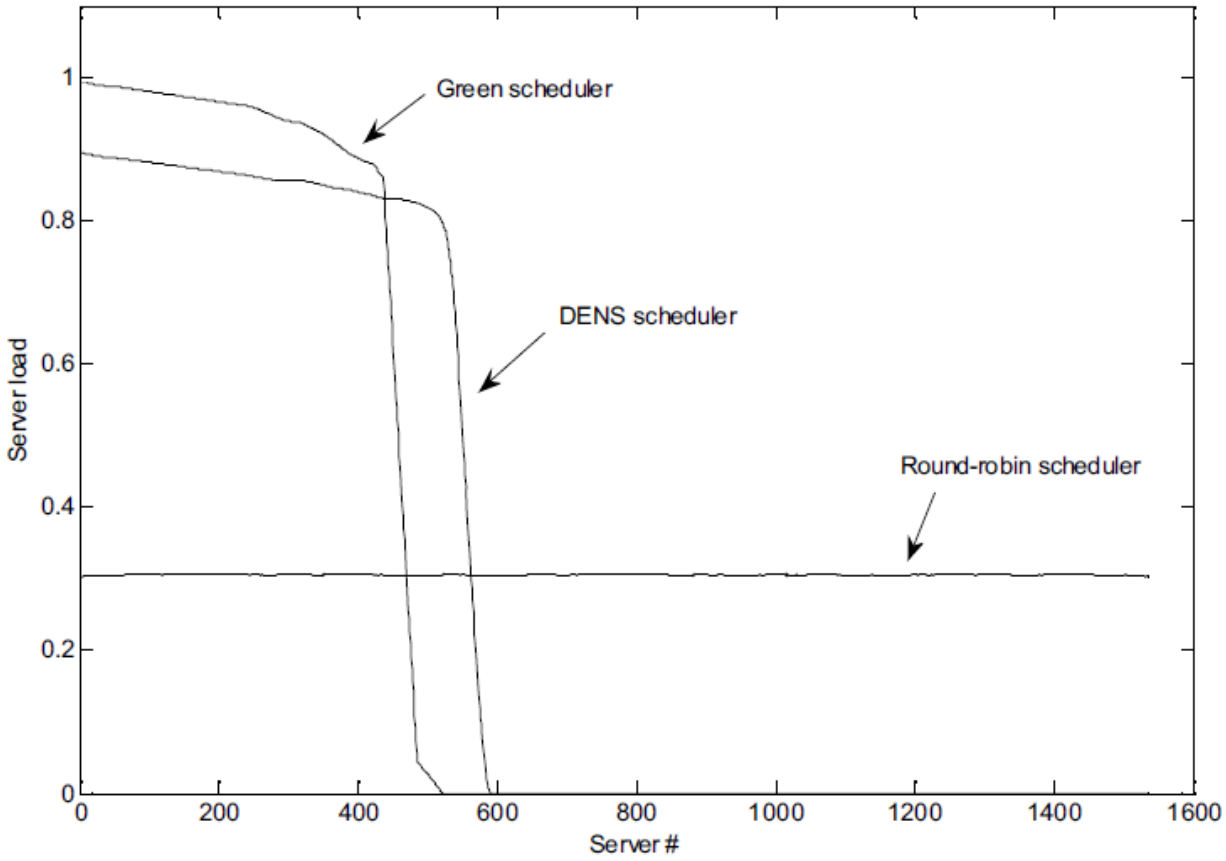


Server selection by DENS metric according to its load and communicational potential

Performance Evaluation(1/5)

- Simulation experiments("GreenCloud" simulator)
  - 1536 servers arranged into 32 racks each holding 48 servers
  - 4 core and 8 aggregation switches
  - Workload generation events are exponentially distributed in time
  - Three evaluated schedulers:
    - DENS scheduler(proposed method)
    - Green scheduler(best-effort workload consolidation on a minimum set of servers)
    - Round-robin scheduler(distributes the workloads equally)

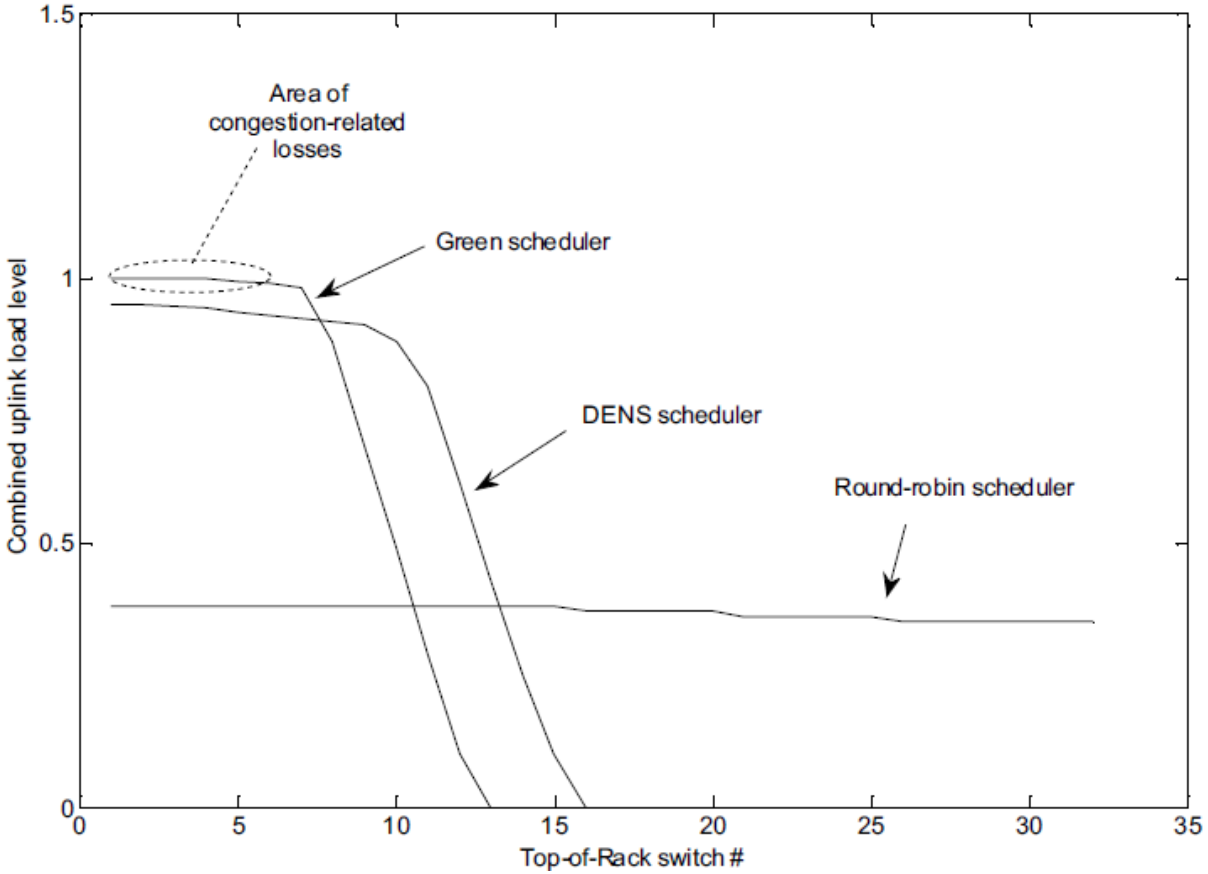
Performance Evaluation(2/5)



Server workload distribution for all three schedulers

Performance Evaluation(3/5)

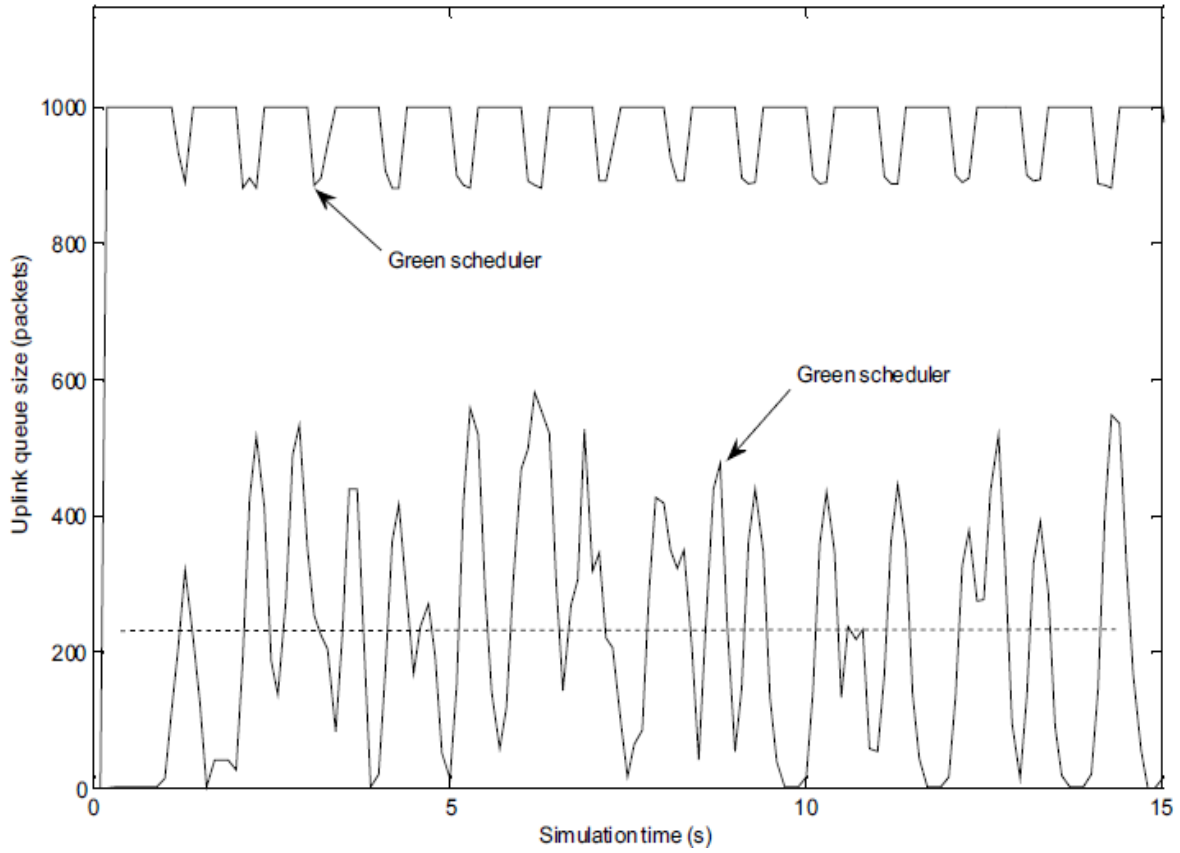
- DENS achieves the workload consolidation for power efficiency while preventing servers and network switches from overloading.



Combined uplink traffic load at the rack switches

Performance Evaluation(4/5)

- Under Green scheduler, the queue remains almost constantly full, which causes multiple congestion losses.
- Under DNS scheduler, the buffer occupancy is mostly below the half of its size with an average of 213 packets.



ToR switch uplink buffer occupancy

Performance Evaluation(5/5)

Data center energy consumption

Parameter	Power Consumption (kW·h)		
	Round Robin scheduler	*Green Scheduler*	*DENS scheduler*
Data center	417.5K	203.3K (48%)	212.1K (50%)
Servers	353.7K	161.8K (45%)	168.2K (47%)
Network switches	63.8K	41.5K (65%)	43.9K (68%)

- Most energy inefficient is a round-robin scheduler .
  - It does not allow any of servers or network switches to be powered down.
- Green scheduler is the most energy efficient.
  - No consideration of network congestion levels.
- DENS uses network awareness to detect congestion hotspots and adjust its job consolidation accordingly.

Conclusions(1/1)

- DENS balances the energy consumption of a data center, individual job performance, and traffic demands.
  - optimizes the tradeoff between job consolidation(to minimize the amount of computing servers)
  - distribution of traffic patterns(to avoid hotspots in the data center network)

Paper Study-3

3) “Improving the Scalability of Data Center Networks with Traffic-aware Virtual Machine Placement”

- Introduction
- Background
- VM placement problem
- Impact of network architectures and traffic patterns on optimal VM placement
- Experiment
- Conclusion

Introduction

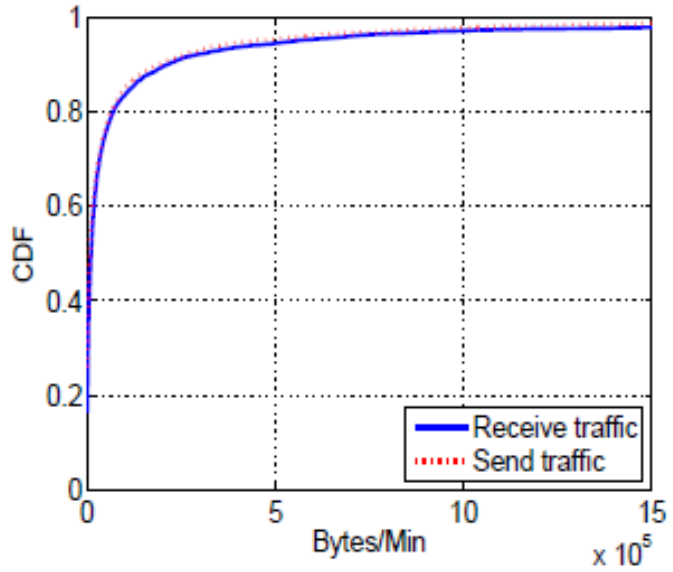
- Normally VM placement is consider for CPU, physical memory and power consumption savings, yet without considering consumption of network resources .

- As a result, this can lead to situations in which VM pairs with heavy traffic among them are placed on host machines with large network cost between them.

- In this paper, we optimize the placement of VMs on host machines.

Background

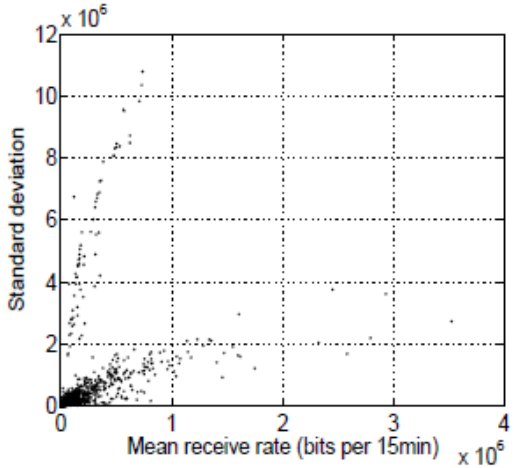
- Uneven distribution of traffic volumes from VMs
  - 80% of VMs have average rate less than 800 KBytes/min, 4% of them have a rate ten times higher.



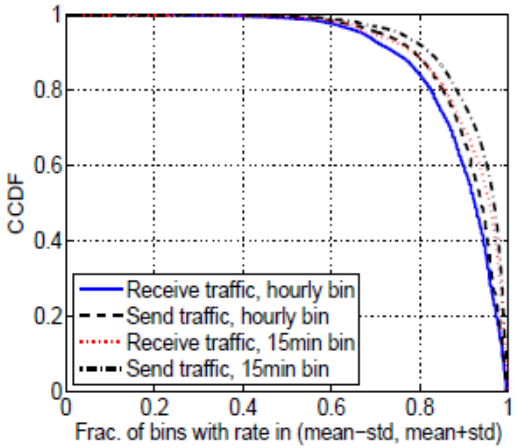
(a) CDF of mean traffic rate

Background

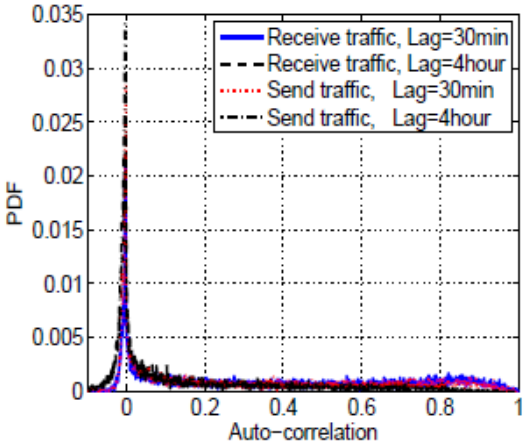
- Stable per-VM traffic at large timescale
  - In Figure (b),show it shows that for the majority of VMs (82%), the standard deviation of their traffic rates is no more than two times of the mean.
  - In Figure (c), show that no less than 80% of the entire set of time intervals are stable.
  - Figure (d),show a large fraction of the VMs' traffic rates are relatively constant.



(b) Distribution of (mean, standard deviation)



(c) CCDF of proportions of stable intervals



(d) PDF of auto-correlation coefficients

Background

- Weak correlation between traffic rate and latency
  - VM pairs with high rate do not necessarily correspond to low latency and vice versa.
  - The correlation coefficient is -0.32

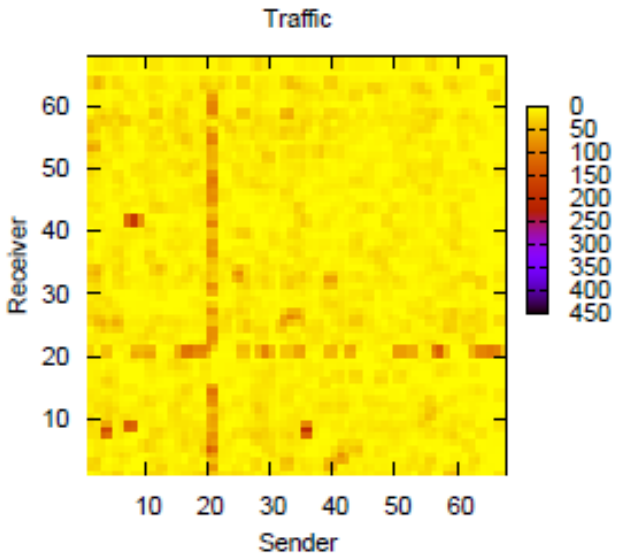


Fig. 2. Traffic matrix

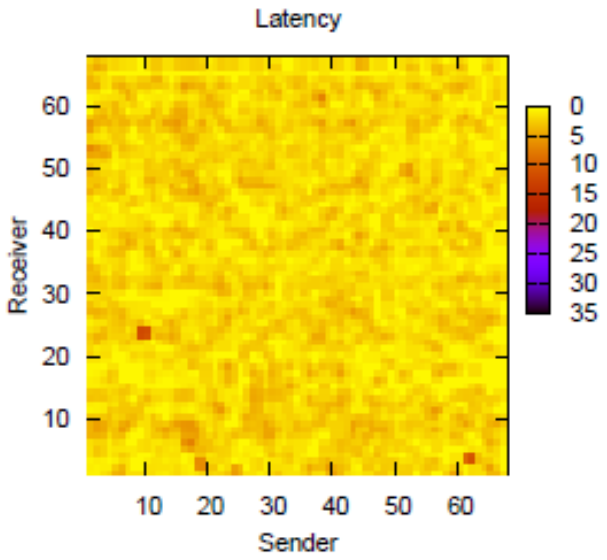


Fig. 3. Latency matrix

Background

- By the above three observation ,for shuffling VM placement such that more traffic are localized and the pairwise traffic demand is better coordinated with the pairwise network cost .

VM placement problem

- We use a ***slot*** *to* refer to one CPU/memory allocation on a host. And assume there are n VMs and n slots.
- We can formally define the Traffic-aware VM Placement Problem (TVMPP) as finding a to <u>minimize</u> the following objective function

$$\sum_{i,j=1,\ldots,n} D_{ij} C_{\pi(i)\pi(j)} + \sum_{i=1,\ldots,n} e_i g_{\pi(i)}$$

C_{ij}= communication cost from slot i to j (we defined as the number of switches on the routing path)

D_{ij}=traffic rate from VM i to j

e_i=external traffic rate for VM i

g_i =the communication cost between VM i and the gateway

Notice: because the cost between every end host and the gateway is the same, the second part can be ignored.

VM placement problem

Proposition 1: [20] Suppose $0 \leq a_1 \leq a_2 \ldots \leq a_n$ and $0 \leq b_1 \leq b_2 \ldots \leq b_n$, the following inequalities hold for any permutation π on $[1, \ldots, n]$

$$\sum_{i=1}^{n} a_i b_{n-i+1} \leq \sum_{i=1}^{n} a_i b_{\pi(i)} \leq \sum_{i=1}^{n} a_i b_i$$

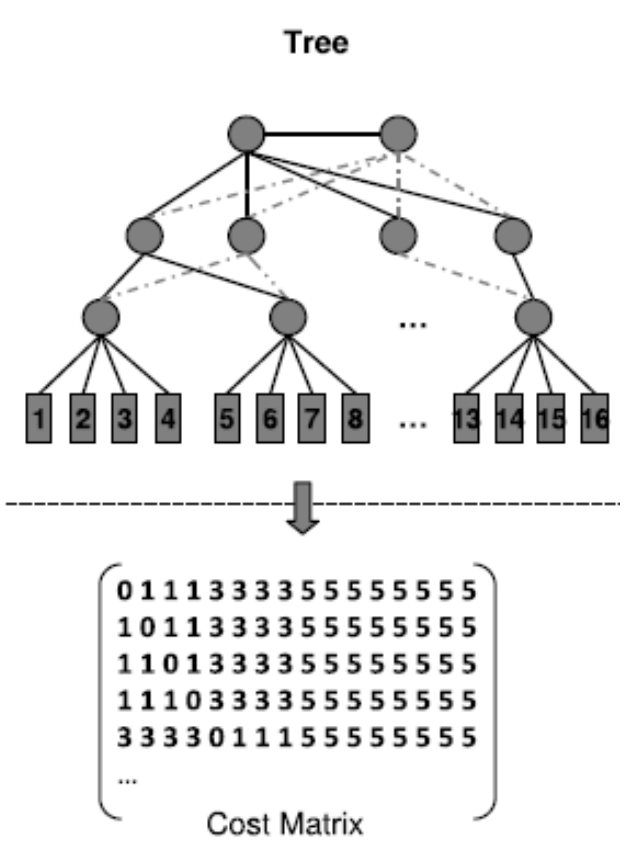
- Base on proposition 1, solving TVMPP is intuitively equivalent to finding a mapping of VMs to slots such that VM pairs with heavy mutual traffic be assigned to slot pairs with low-cost connections.

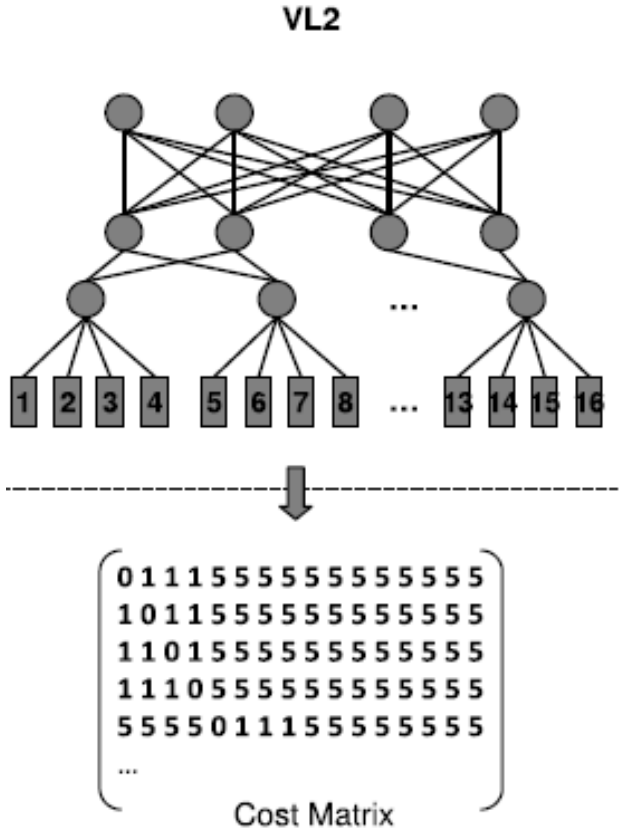
VM placement problem

- Algorithms: *Cluster-and-Cut*
  - map each VM-cluster to a slot-cluster
    - VM-clusters are obtained via classical min-cut graph algorithm which ensures that VM pairs with high mutual traffic rate are within the same VM-cluster.
    - Slot-clusters are obtained via standard clustering techniques which ensures slot pairs with low-cost connections belong to the same slot-cluster.
    - Solving it recursively until the size of cluster is small.
  - VM pairs with heavy mutual traffic have be assigned to slot pairs with low-cost connections.

Impact of network architectures and traffic patterns on optimal VM placement

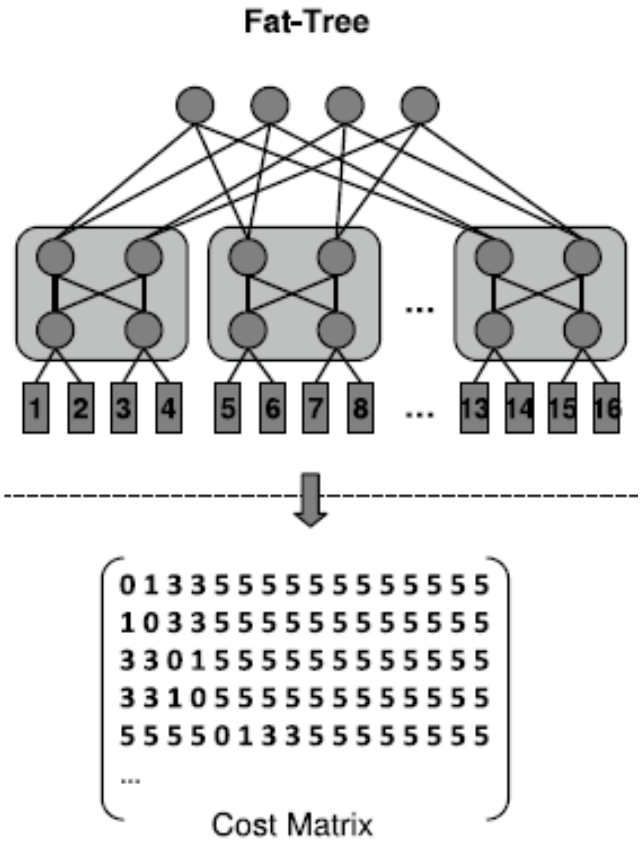
- Four data center network architectures
  - **Tree**
  - **VL2**



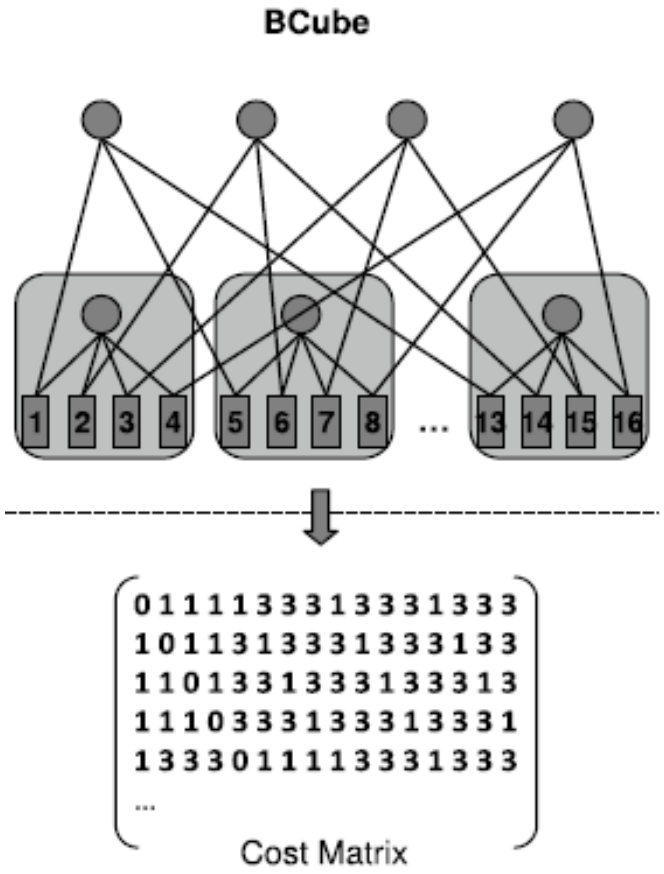


Impact of network architectures and traffic patterns on optimal VM placement

- **Fat-Tree**



- **BCube**



Impact of network architectures and traffic patterns on optimal VM placement

- Two traffic patterns
  - global traffic model

    each VM communicates with every other at a constant rate
  - partitioned traffic model

    only VMs within the same partition communicate with each other

  - Traffic patterns in reality can be roughly considered to be generated from a mixture of these two special models.

Impact of network architectures and traffic patterns on optimal VM placement

- In *Global Traffic Model*
  - BCube has the largest improvement space, the VL2 comes with the smallest.
  - Because BCube network has four levels of intermediate switches which has larger path cost. Thus , optimizing VM placement a higher gain.
  - Due to VL2 is load balancing so TVMPP is no effect.

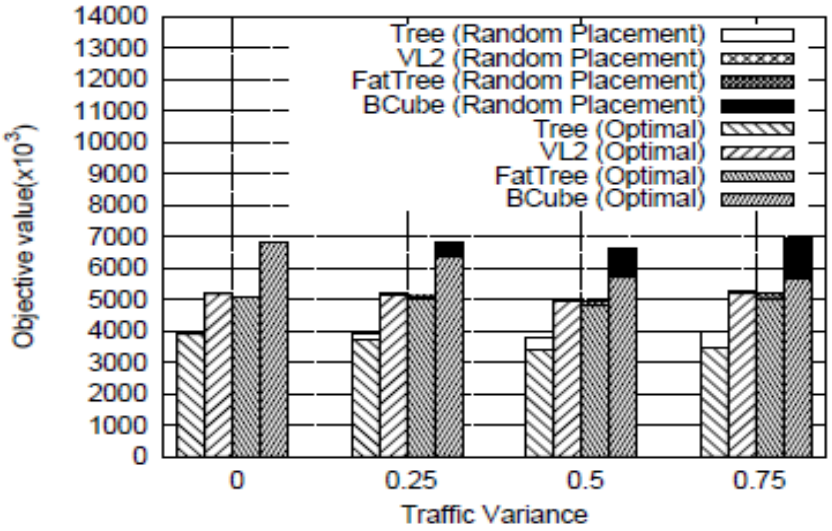


Fig. 5. Optimal objective value *vs* objective value achieved by random placement (global traffic model, different traffic variance)

Impact of network architectures and traffic patterns on optimal VM placement

- In *Partitioned Traffic Model*
  - GLB is the low bound of object value.
  - Under the Bcube, the gap is larger.

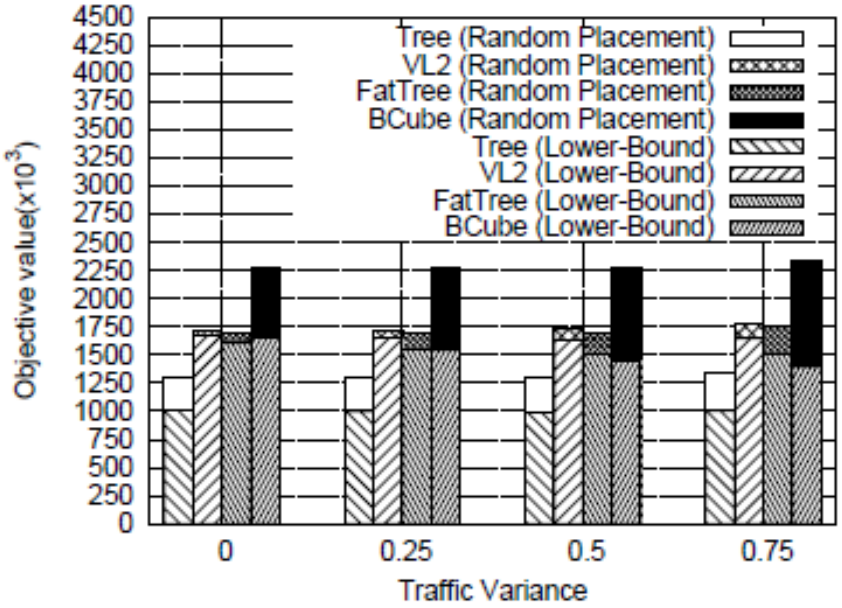


Fig. 6. GLB *vs* objective value achieved by random placement (partitioned traffic model, 16 partitions of 64 VMs each)

Impact of network architectures and traffic patterns on optimal VM placement

- different partition size
  - the improvement potential is proportionally higher for smaller partition sizes.

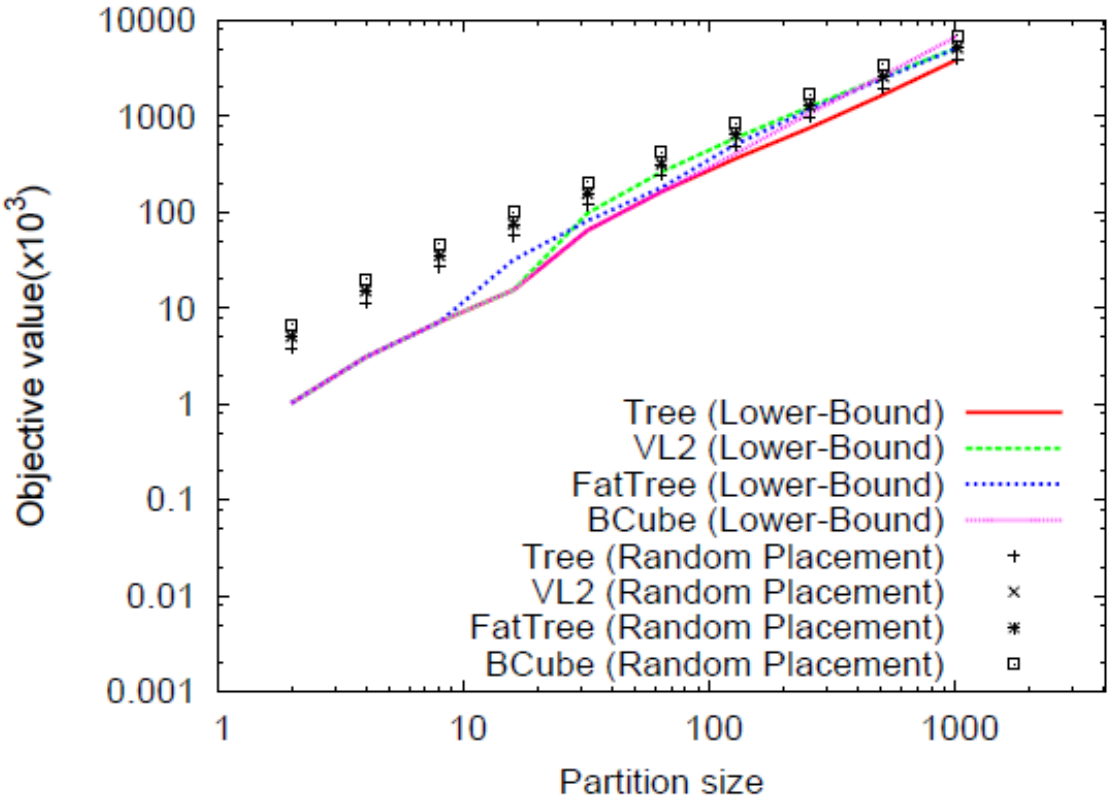


Fig. 8. GLB *vs* objective value of random placement (with different partition size)

Experiment

- The objective function value given by the Cluster-and-Cut is about 10% smaller than the two benchmarks, with CPU time being halved.
- Cluster-and-Cut algorithm is effective.

Topology	Algorithms	Gilmore-Lawler bound	Performance	
			Best value	CPU min
Tree	LOPI	4.63e+10	8.22e+10	22
	SA		8.35e+10	27
	Cluster-and-Cut		8.13e+10	11
VL2	LOPI	7.03e+10	1.09e+11	25
	SA		1.12e+11	31
	Cluster-and-Cut		1.05e+11	12
Fat-tree	LOPI	6.43e+10	1.07e+11	26
	SA		1.12e+11	32
	Cluster-and-Cut		0.97e+11	13
BCube	LOPI	5.55e+10	1.43e+11	29
	SA		1.41e+11	35
	Cluster-and-Cut		1.21e+11	14

TABLE I
ALGORITHM PERFORMANCE WITH HYBRID TRAFFIC MODEL

Conclusion

- A careful VM placement can localize large chunks of traffic and thus reduce load at high-level switches.

Question & Answer

- Q:
  - Energy consumption is a growing concern for data centers operators, in this issue what is the main energy consumption in the Data Center? And what are the two main techniques to solve above problem?
- A:
  - 1.Computing servers account for a major portion of data center energy consumption.
  - 2.Dynamic Voltage and Frequency Scaling (DVFS) and Dynamic Power Management (DPM)