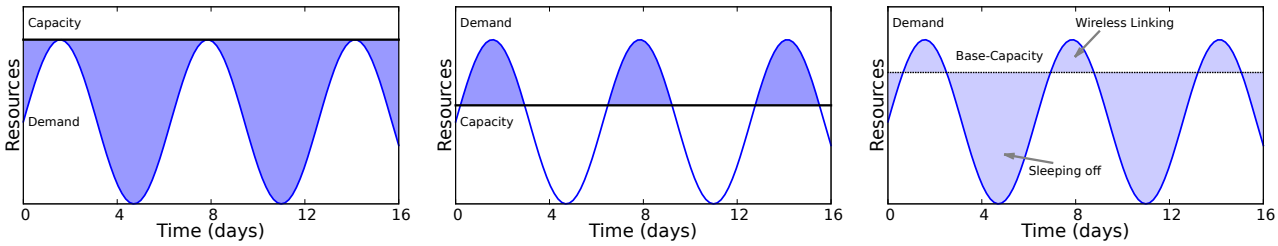
Report on Data Centers

Yongsen MA

I. Motivation

With the increasing demand of traffic loads and user requirements, it is a great challenge to make data centers agile and energy-efficient. A basic solution is to allow dynamic resource allocation based on flexible data center networks. The recent development of 60GHz wireless technology opens a door for the deployment of flexible data centers. It provides a good choice of adding capacity to data centers with under-provisioning capacity design. However, this will lead to new problems such as topology control and capacity scheduling, apart from the wireless propagation and link adaption in wireless networks. The PHY and MAC standardization of 60GHz networks is still ongoing (WirelessHD and IEEE 802.11ad/WiGig), its application in data centers should be further explored for the specific feature of topology and services.



(a) Provisioning for peak load (b) Fixed underprovisioning (c) Flexible underprovisioning

Fig. 1. Capacity provisioning based on traffic loads

Generally, the capacity of data centers is designed for peak loads, as shown in Figure 1a, but the resources will be wasted during non-peak times [Armbrust et al., 2010]. On the other hand, if the capacity is designed in under-provisioning case as shown in Figure 1b, the peak requirements can not be satisfied, leading to potential revenue sacrifice or over-occupied errors. To address these problems, we can employ flexible capacity and dynamic demand response, i.e., 60GHz wireless linking to add capacity for peak loads and sleeping mechanism to improve performance/cost efficiency, as shown in Figure 1c.

The flexible data centers are generally composed of:

1) Topology Control:
   a) Conflict Graph: Propagation table (wireless), capacity matrix (wired/wireless) and traffic loads table (wired/wireless)
   b) Topology Graph: Physical-specific ID Addresses (PAs), Location-specific IP Addresses (LAs) and Application-specific IP Addresses (AAs)
2) Addressing:
   a) PAs: providing device information for Topology Control
   b) LAs: providing distance information for 3D Beamforming
   c) AAs: providing traffic information for Demand Response
3) Capacity Scheduling:
   a) 3D Beamforming: adding additional capacity to neighbor ToRs according to LAs
   b) Demand Response: traffic estimation according to AAs
   c) Link Adaption: capacity scheduling according to Conflict Graph and Topology Graph

II. Methodology

A. Measurement

Measurement:

- Propagation table
- Capacity matrix
- Traffic loads

Topology changes:

- Errors: nodes, links, miswiring
- wireless linking
- Sleeping on-off

Mapping:

- PA to LA (Beamforming)
- LA to AA (Traffic estimation)

Connections:

- wired 10G
- wireless direct 6G
- wireless indirect 1/2G

wireless data centers: making data centers flexible, i.e., using topology control to make resource allocation responding to traffic patterns and requirements.

- graph mapping: topology changes due to errors or wireless links.
- demand response: topology and capacity (wired and wireless) are stable for certain graphs, so the problem is demand estimation.

So how to allocate wireless links (direct links between neighbor nodes or 3D beamforming for remote nodes), to deal with the problem of demand response when topology graphs are changing. Topology changes may occur due to errors, sleeping mechanism or wireless links.

reasons for compress sensing:

1) topology graphs are **large**: the requirement for compress sensing to reduce mapping overhead
2) wireless propagation and traffic demand measurement is challenging: requirements
3) topology graphs are **sparse**: the prior condition for compress sensing
4) topology graphs have **local changes** (errors, on-off or wireless) in real-time operating: conditions

B. Mapping

C. Scheduling

III. Evaluation

IV. Literature Review

"Generic and Automatic Address Configuration for Data Center Networks", SIGCOMM 2010

Basic Procedures:

1) O2 Mapping
   a) Candidate selection via SPLD: **select** candidate with the same SPLD.
   b) Candidate filtering via orbit: **skip** candidate with the same orbit, then $Decomposition()$.
   c) Selective splitting $Refinement^*()$: **split** cells that really connect to the including cell.
2) Malfunction Detection
   a) Anchor pair selection:
   b) Malfunction detection:

Problems:

1) Initial selection of vertex $\nu \in \pi_p^i$
2) Compute complexity of O2
3) Whether it can be resolved by Compress Sensing?
   a) the topology graphs are sparse.
   b) only certain parts are changing in real-time operating (considering certain servers can be turned down for energy-efficiency and demand response).

"OSA: An Optical Switching Architecture for Data Center Networks with Unprecedented Flexibility", NSDI 2012

Architecture

"BCube: A High Performance, Server-centric Network Architecture for Modular Data Centers", SIGCOMM 2009

"VL2: A Scalable and Flexible Data Center Network", SIGCOMM 2009

1. Background:

- Limited server-to-server capacity: over-subscription ratio increases rapidly
- Fragmentation of resources: high turnaround time for reconfiguration
- Poor reliability and utilization: multiple paths waste at most 50% of maximum utilization

2. Measurement:

- Flow: VLB will perform well on this traffic
  - Distribution of flow sizes: majority of flows are small
  - Number of concurrent flows: 10 concurrent flows at more than 50% of the time, 80 concurrent flows at least 5% of the time
- Traffic: it is unlikely that other routing strategies will outperform VLB
  - Poor summarizability of traffic patterns
  - Instability of traffic patterns
- Failure

3. Solution: Virtual Layer 2 Networking

- Topology: D_I Int. Switches and D_A Agg. Switches
- Addressing: Location-specific IP Addresses and Application-specific IP Addresses
- Directory: lookups and updates of AA-to-LA mapping, reactive cache update

Data center traffic analysis:

- arrive intervals.
- application types.

"DCell: A Scalable and Fault-Tolerant Network Structure for Data Centers", SIGCOMM 2008

"A Scalable, Commodity Data Center Network Architecture", SIGCOMM 2008

Topology: k pods, two layers of k/2 switches, each k-port switch connected to k/2 hosts, each remaining k/2 ports connected to k/2 of the k ports in the aggregation layer of the hierarchy. Max $48 \cdot \frac{48}{2} \cdot \frac{48}{2} = 27648$ hosts for 48-port GigE switches.

Addressing: Pod: 10.pod.switch.1, core: 10.k.j.i, host: 10.pod.switch.ID

Wireless

"Mirror Mirror on the Ceiling: Flexible Wireless Links for Data Centers ", SIGCOMM 2012

Wireless Data Center is a new concept which should be explored further. Many issues such as interference, secluding, security, etc. should be standardized. On the other hand, the issues on cost, performance, energy-efficiency, reliability, etc. should be taken into consideration compared with electrical or optical data centers.

1) electrical
2) optical
3) wireless
    a) Link Blockage
    b) Radio Interference

Problem:

1) Apart from the concurrent links, the efficient throughput should be explored. For instance, although the concurrent links are more with larger ceiling height h, it can decrease the throughput according to the curves of RSS (or Data Rate) vs. distance as shown in Figure 5.
2)
3)

Max concurrent links: Link conflicts (SINR); Greedy scheduling (graph coloring); Assigning radios.

"Augmenting Data Center Networks with Multi-gigabit Wireless Links", SIGCOMM 2011

"The base wired network is provisioned for the **average case** and can be oversubscribed. Each ToR switch is equipped withe one or more 60GHz wireless devices." "A central controller monitors DC **traffic patterns**, and switches the beams of the wireless devices to set up flyways between ToR switches that provide **added bandwidth** as needed." So it is ideal for **flexible and energy-efficient** data centers.

Problem:

- Conflict graph: for N racks and K antenna orientations, the input table is very **large** with the size of $(NK)^2$, and the propagation conditions are similar, i.e., the table is **sparse**.

For flexible data centers, 60GHz wireless is an attractive choice for its simplify and inexpensive features. On the other hand, 60GHz wireless is a active research area, such as IEEE 802.11ad, WiGig and WirelessHD standards, that are still under exploring on protocols and technology. For data centers, it should be further explored according to the topology and requirements.

Resource Allocation

"FairCloud: Sharing The Network In Cloud Computing", SIGCOMM 2012

"NetPilot: Automating Datacenter Network Failure Mitigation", SIGCOMM 2012

"Towards Predictable Datacenter Networks", SIGCOMM 2011

REFERENCES

[Armbrust et al., 2010] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A view of cloud computing. *Commun. ACM*, 53(4):50–58.