SIGCOMM Preview Session: Datacenter Networking (DCN)

Hakim Weatherspoon

Associate Professor, Cornell University

SIGCOMM
Florianópolis, Brazil
August 22, 2016

Many slides borrowed from George Porter's Preview session in SIGCOMM 2015

Cloud Computing

- The promise of the Cloud
  - A computer utility; a commodity
  - Catalyst for technology economy
  - Revolutionizing for health care, financial systems, scientific research, and society

Cloud Computing

- The promise of the Cloud
  - *ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*

NIST Cloud Definition

Google Compute Engine

amazon web services

Windows Azure

Microsoft Office 365

Google

salesforce

amazon

NETFLIX

facebook

Spotify

Cloud Computing

- The promise of the Cloud
  - *ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*

NIST Cloud Definition

Microsoft® Office 365

Google

salesforce

Google Compute Engine

amazon web services™

amazon

NETFLIX

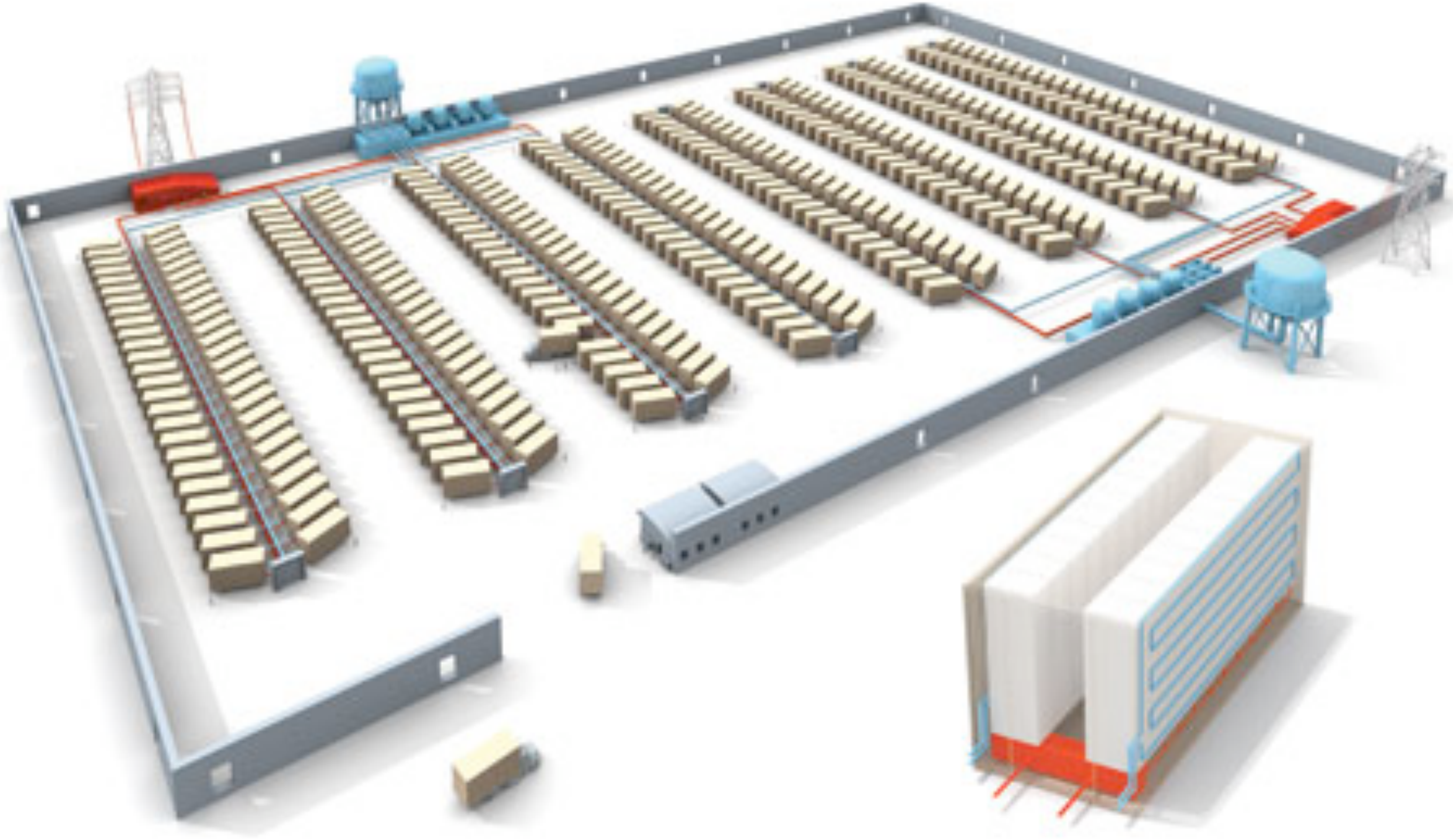
Windows Azure

facebook

Spotify®

Cloud Computing needs Datacenters

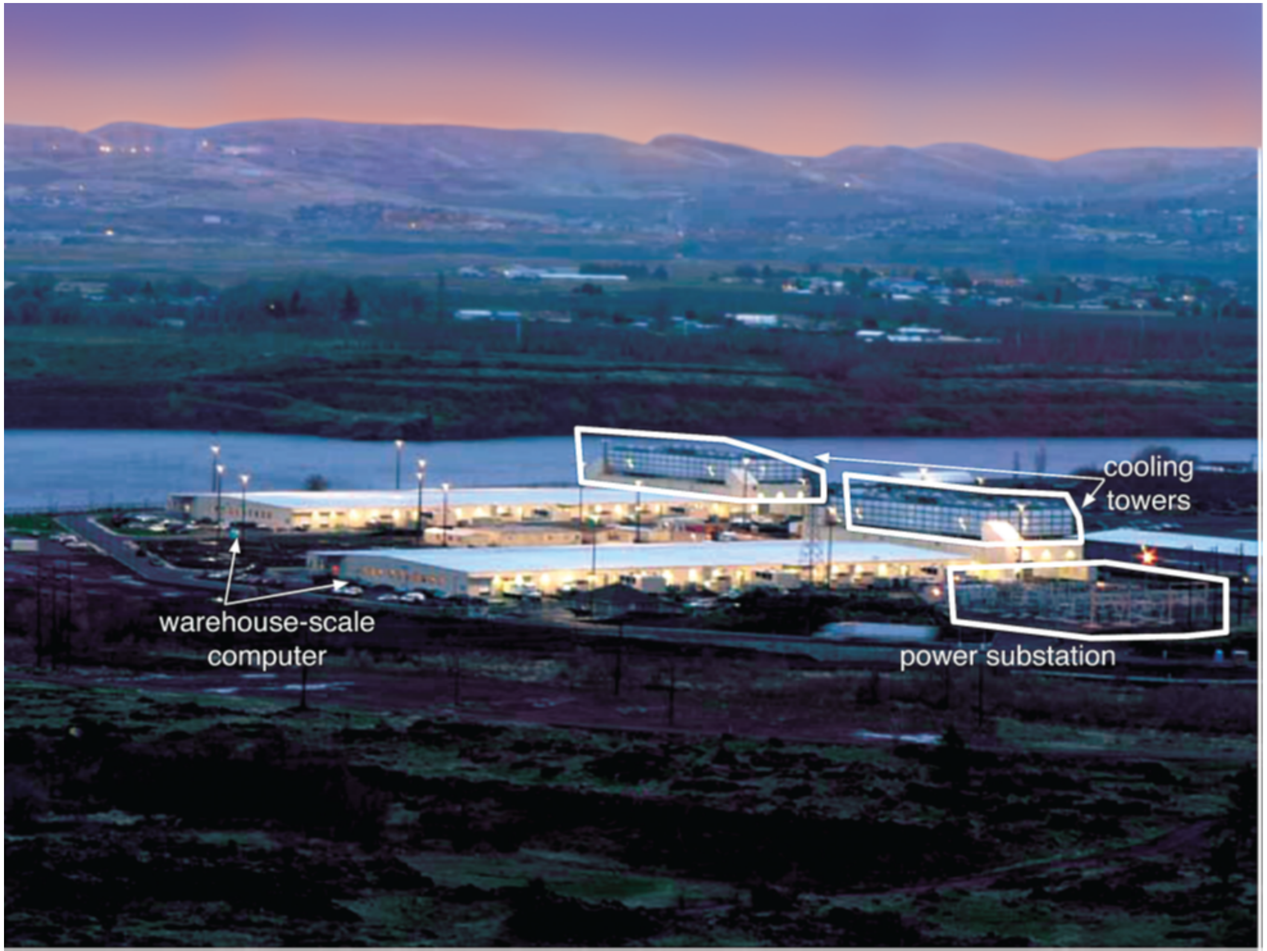
- How big is the Cloud?
  - Exabytes: Delivery of petabytes of storage daily



Titan tech boom, randy katz, 2008

Cloud Computing needs Datacenters

- How big is the Cloud?
  - Most of the worlds data (and computation) hosted by few companies in datacenters



Cloud Computing needs Datacenters

- How big is the Cloud?
  - Most of the worlds data (and computation) hosted by few companies in datacenters

Inside of a Datacenter

- 10s to 100s of thousands of servers
- Exabytes (1000s of petabytes) of storage
- Infrastructure-as-a-Service (IaaS)
  - Amazon EC2, Google Compute Engine, Microsoft Azure
- Single "application" spread across many thousands of servers (e.g. Amazon.com)
  - Application components such as caches, web servers, data bases, distributed file servers,…
  - Each component is "scaled" to meet the needs of millions (or billions) of users

Why Study DCN

- Scale
  - Google: 0 to 1B users in ~15 years
  - Facebook: 0 to 1B users in ~10 Years
  - *Must operate at the scale of O(1M+) users*
- Cost:
  - To build: Google ($3B/year), MSFT ($15B/total)
  - To operate: 1---2% of global energy consumption*
  - *Must deliver apps using efficient HW/SW footprint*

* LBNL, 2013

What defines a datacenter network (DCN)

The Internet	Datacenter Network (DCN)
Many autonomous systems (ASes)	One administrative domain
Distributed Control/routing	Centralized Control and route selection
Single shortest path routing	Many paths from source to destination
Hard to measure	Easy to measure
Standardized Transport (TCP and UDP)	Many transports (DCTCP, qFabric,…)
Innovation requires consensus (IETF)	Single company can innovate
“Network of networks”	“Backplane of giant supercomputer”

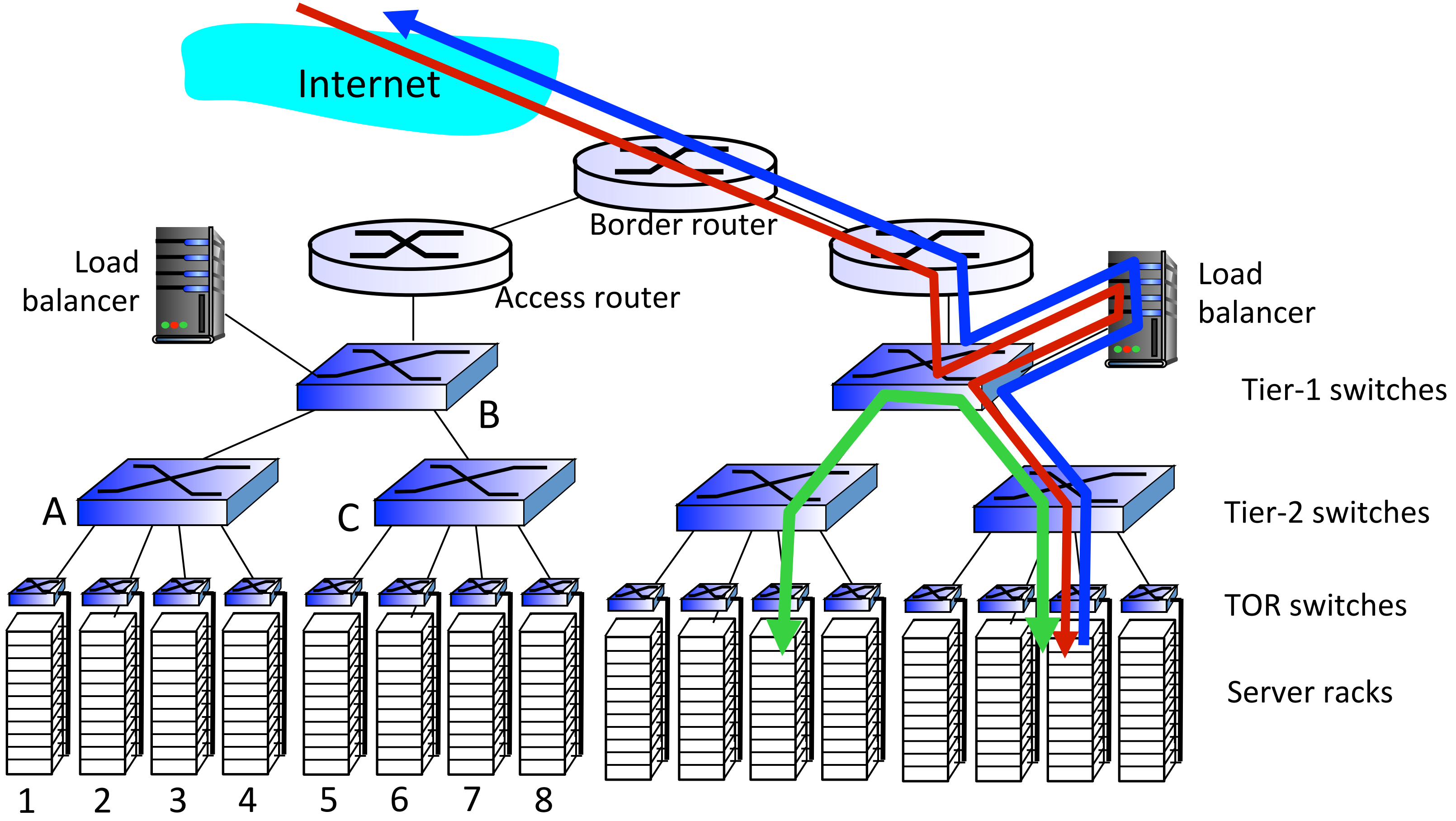
DCN Research “cheat sheet”

- How would you design a network to support 1M endpoints?
- If you could…
  - Control all the endpoints and the network?
  - Violate layering, end---to---end principle?
  - Build custom hardware?
  - Assume common OS, Dataplane functions?

- Top-to-bottom rethinking of the network

DCN Topologies: Background

DCN Topologies: Tree-based



DCN Topologies: Tree-based

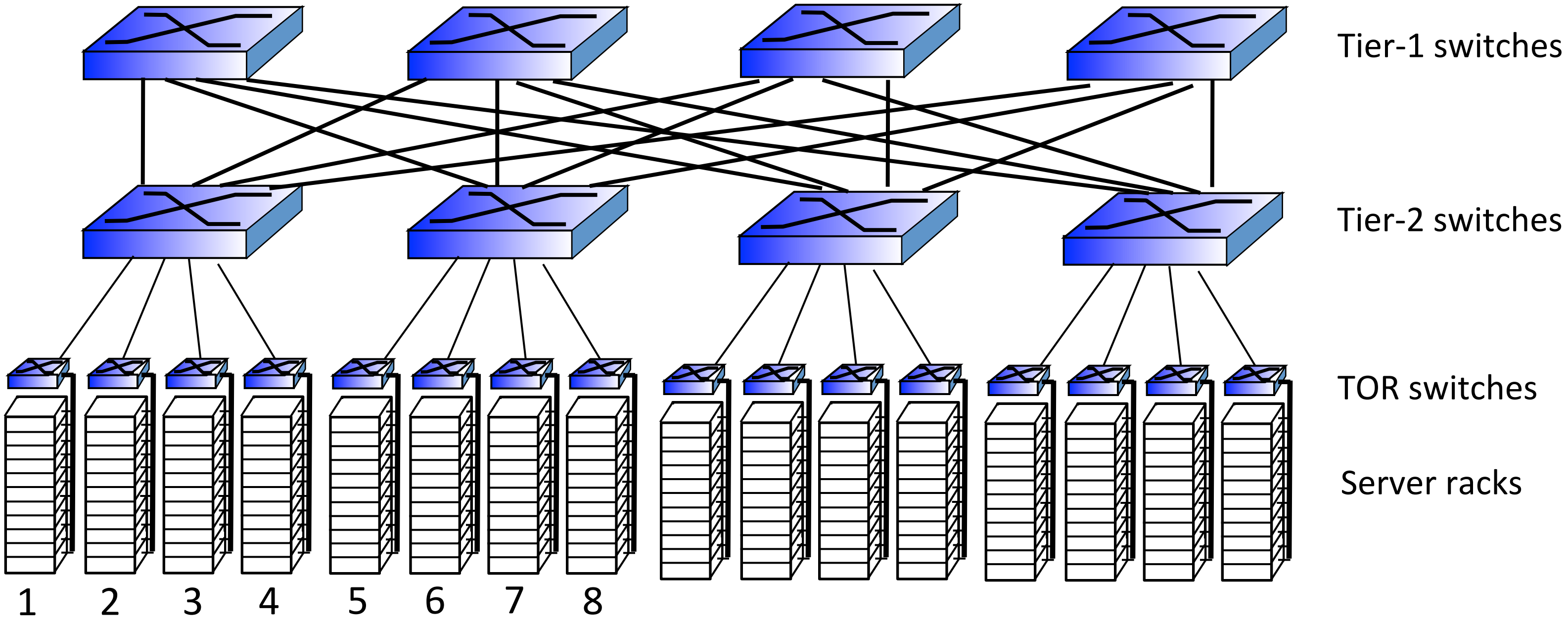
Issues with Traditional Data Center Topology

- *Oversubscription:*
  - Ratio of the worst-case achievable aggregate bandwidth among the end hosts to the total bisection bandwidth of a particular communication topology
  - Lower the total cost of the design
  - Typical designs: factor of 2:5:1 (400 Mbps)to 8:1(125 Mbps)
- *Cost:*
  - Edge: $7,000 for each 48-port GigE switch
  - Aggregation and core: $700,000 for 128-port 10GigE switches
  - Cabling costs are not considered!

DCN Topologies: Folded-Clos multi-rooted tree

“FatTree” overcomes limitations

- rich interconnection among switches, racks:
  - increased throughput between racks (multiple routing paths possible)
  - increased reliability via redundancy



Paper Previews: Datacenters 1

- Session, Wednesday, 10:40am to 12:20pm
- Globally Synchronized Time via datacenter networks
  - Ki Suh Lee, Han Wang, Vishal Shrivastav
  - Provides tight and bounded precision for synchronizing time throughout an entire datacenter.

Paper Previews: Datacenters 1

- Session, Wednesday, 10:40am to 12:20pm
- Robotron: Top-down network management at Facebook
  - Yu-Wei Eric Sung, Xiaozheng Tie, Startsky H.Y. Wong, Hongyi Zeng
  - Network management via separating intent (expressed by Engineers) from implementation (translated by the system), which making the system more robust. Further, Robotron monitors operational state to ensure conformance to desired state.

Paper Previews: Datacenters 1

- Session, Wednesday, 10:40am to 12:20pm
- RDMA over Commodity Ethernet at Scale
  - Haitao Wu, Zhong Deng, Gaurav Soni, Jianxi Ye, Jitu Padhye, Marina Lipshteyn
  - Challenges and approaches to using RDMA over commodity Ethernet (RoCEv2). Paper shows that RoCEv2 scale and issues can be addressed and that RDMA can replace TCP in the datacenter.

Paper Previews: Datacenters 1

- Session, Wednesday, 10:40am to 12:20pm
- ProjecToR: Agile Reconfigurable Data Center Interconnect
  - Monia Ghobadi, Ratul Mahajan, Amar Phanishayee, Nikhil Devanur, Janardhan Kulkarni, Gireeja Ranade, Pierre-Alexandre Blanche, Houman Rastegarfar, Madeleine Glick, Daniel Kilper
  - Explores use of free-space optics for building datacenter interconnects using digital micromirror devices (DMD) and mirror assembly combination as a transmitter and photodetector on top of the rack as a receiver.

Final thoughts

- DCN Is an exciting, fun research area
- While many papers are from Microsoft, Google, Facebook, …
  - YOU have the ability to have enormous impact
  - Many Projects are open---source
  - E.g., http://sonic.cs.cornell.edu
- Rethink the entire network stack!
  - Hardware, software, protocols, OS, NIC, …