Cutting the Cord: A Robust Wireless Facilities Network for Data Centers

Yibo Zhu, Xia Zhou[§], Zengbin Zhang, Lin Zhou,
Amin Vahdat[†], Ben Y. Zhao and Haitao Zheng

U.C. Santa Barbara, [§]Dartmouth College,
[†]U.C. San Diego & Google
yibo@cs.ucsb.edu

Data Center Networks (DCN)

- DCN: key infrastructures for mobile and big data applications

Google Walmart JPMorgan

- Large and dynamic → management complexity
  - Highly dynamic data traffic
  - Shared by changing customers
  - Frequent failure, maintenance and upgrades

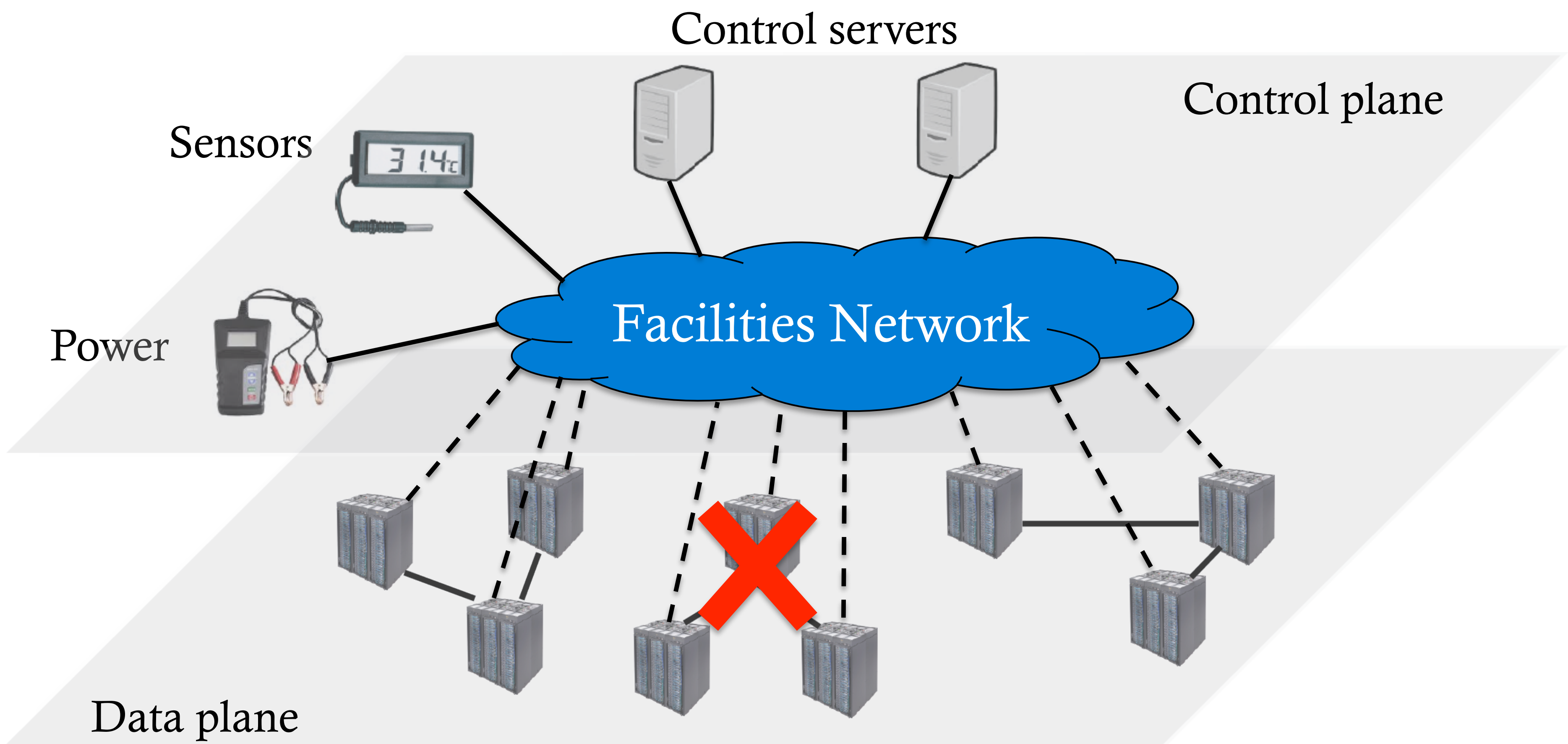
Beyond Data Plane

- Various control messages
  - Flow scheduling
  - Monitoring environment & power
  - Virtual machine imaging and configuration
  - Failure recovery
  - Bootstrap upgraded devices

- Must deliver timely and reliable
  - Not interfered by congested data traffic
  - Even when data plane not working

Upgrade ~100 servers per day on average

A Facilities Network



Proposed DCN architecture

Requirements of Facilities Network

Performance

<u>Low bandwidth</u>

- 1Gbps enough

<u>Bounded delay[1]</u>

- One packet message <10ms
- 1MB Large message <500ms

[1]*Devoflow*, SIGCOMM'11

Fault isolation

<u>Not fate-sharing</u>

- Ideally physically separated

Robustness

<u>Always connected</u>

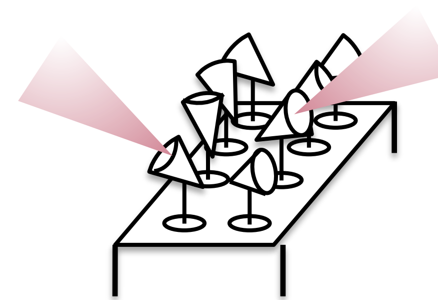
- Even when large portions down

Must remain working even racks taken off

Option: Wired Facilities Network

- Connect all devices using cables
  - In-band: share w/ data plane
  - Out-of-band
- Advantage: large capacity
- Challenges
  - Out-of-band: high cost, wiring headache
  - Poor fault isolation/robustness
    - Zero fault isolation for in-band
    - Even out-of-band interrupted by cable tray maintenance

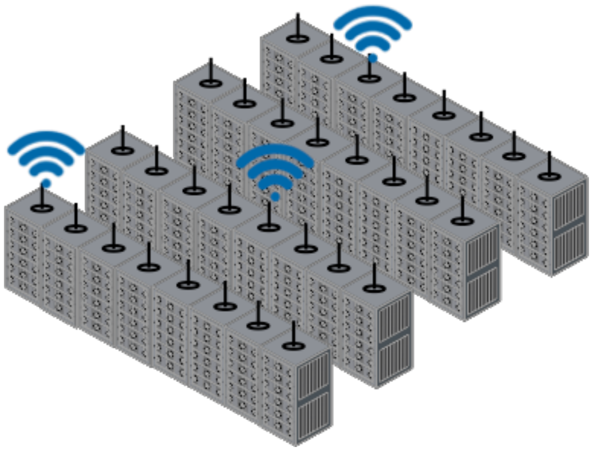
Option: Wireless Facilities Network



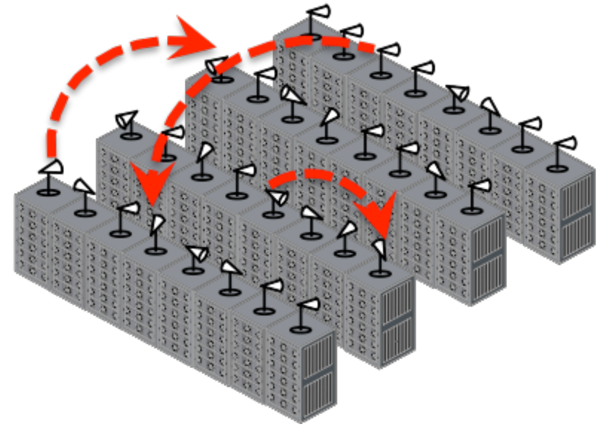
- Add radios to racks
  - WiFi (1.3Gbps), 60GHz (6.76Gbps)
  - Enough bandwidth
- Advantages
  - Cost: low (no additional switches/cables)
  - Fault isolation: physically isolated from data plane
  - Robustness: automatically reform links
- Challenge: delay from wireless interference

Choice of Wireless Technology

WiFi



60GHz 3D Beamforming



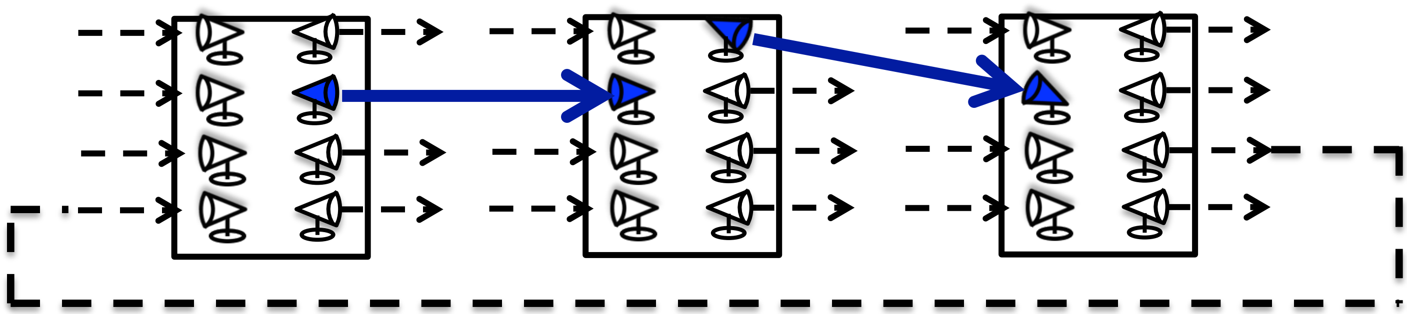
WiFi	60GHz 3D Beamforming
Widely available	Recently available
Well-understood • Omni-directional • Contend for channel	Less-understood • Highly directional • Need coordination
Large interference footprint • Poor in dense DC • Unpredictable delay	Small interference footprint • Good for dense DC

Outline

- Motivation
- System design
  - Angora: a 60GHz facilities network
  - Wireless overlay design
  - Minimizing link interference
  - Fault recovery
- Evaluation
- Conclusion

Angora: a 60GHz Overlay

- Highly directional signal + limited radios per rack → limited connections per rack
- Antenna alignment → extra delay



- Angora: fixed topology overlay
  - Multi-hop → any-to-any connectivity
  - Fixed topology → no link coordination → no extra controllers, minimize delay

Structured Overlay Graph

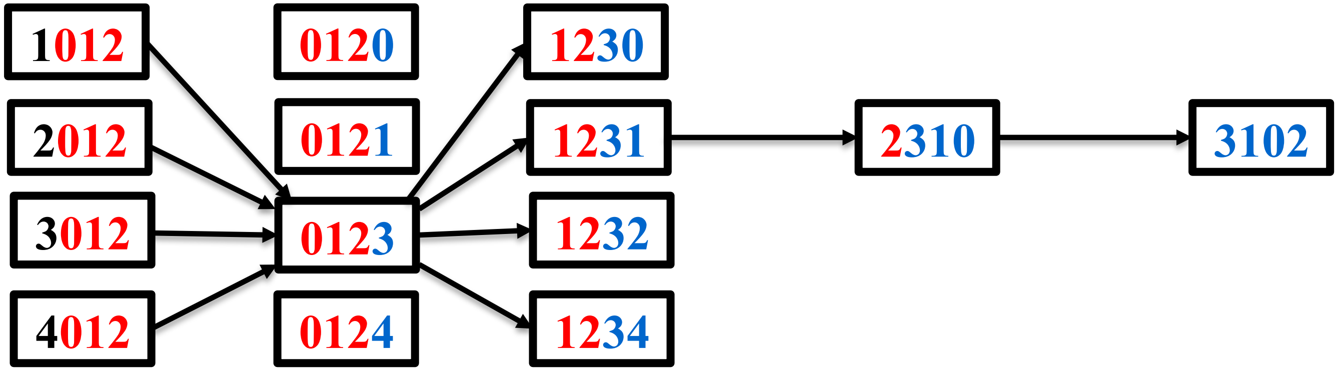
- Key goal: minimize delay (hop count)
- The constraint: constant number of radios per rack → constant degree graph

- We choose **Kautz** graph
  - Smallest diameter given node degree and the number of nodes.
- Hop count: Kautz < Random[1] << Fat-tree
  - Wired networks prefer Fat-tree due to low wiring complexity

[1] *Jellyfish*, NSDI'12

Kautz Graph

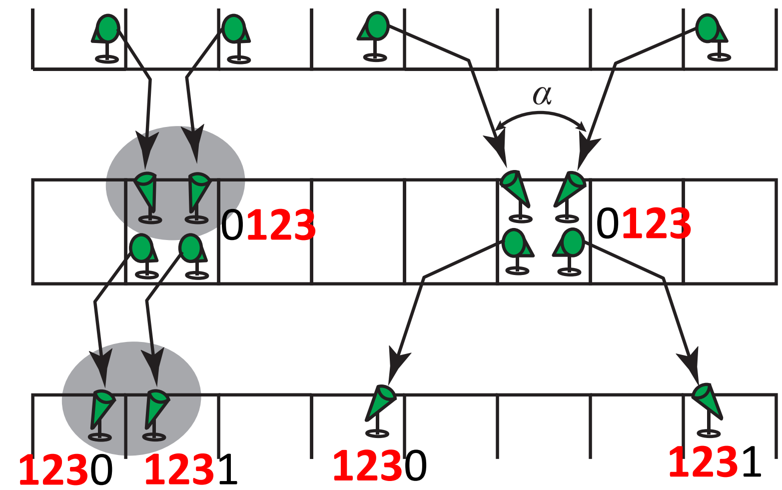
- Simple digit-shift routing



- Graph diameter = length of IDs = ~$\log_k(N)$
  - N: # of nodes, k: node degree (4)
- Challenge: Kautz only supports specific N
  - We design an algorithm to handle arbitrary N

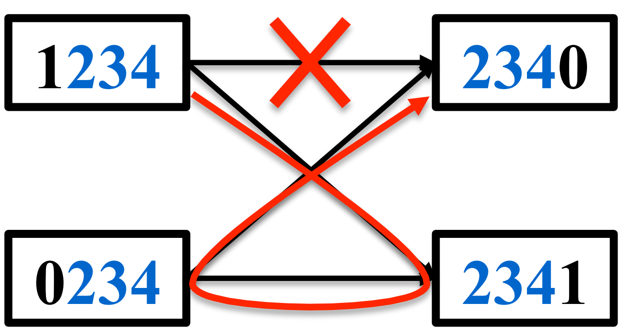
Node Naming and Interference

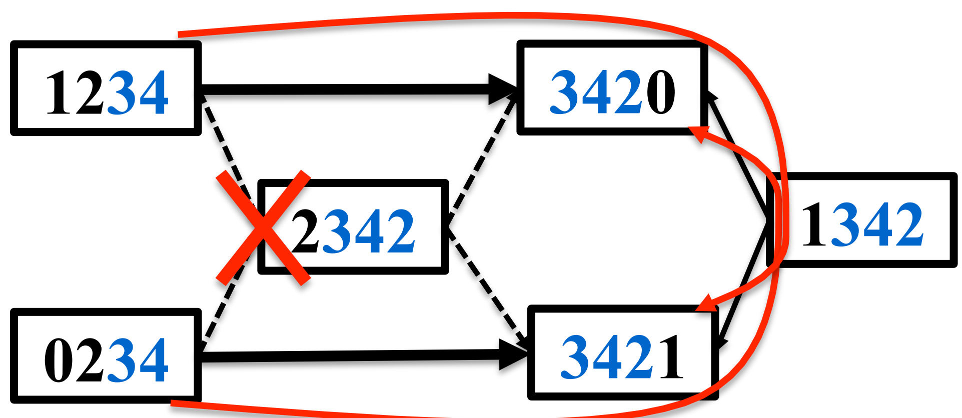
- Nodes naming affects interference
  - 60GHz interference: function of angular separation
- Goal: maximize angular separation between links
- Designed an optimal naming scheme
  - Achieved 14° angular separation in practice



Failure Recovery Algorithms

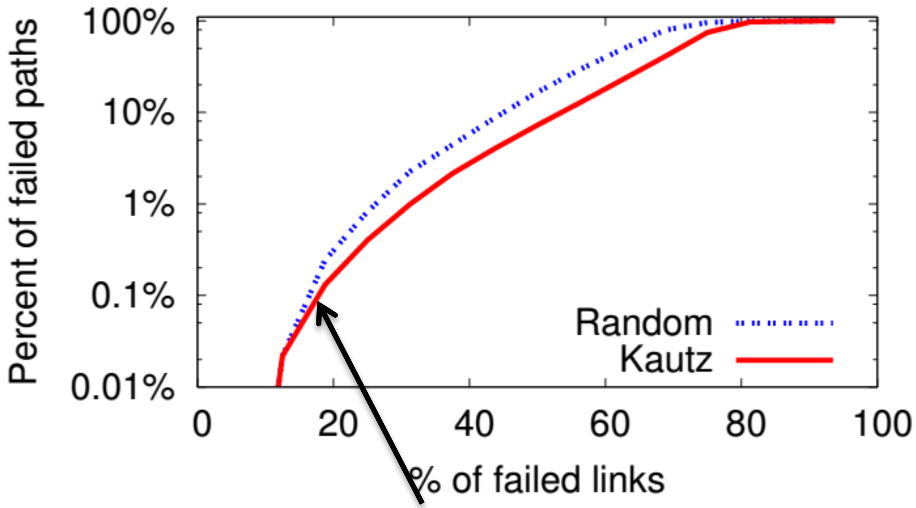
- Link failure → remove a graph edge
  - May happen when radio fails, or signal blocked
  - Leverage Kautz structure to re-route the traffic
- Rack failure → remove a graph node
  - Similar deterministic algorithm





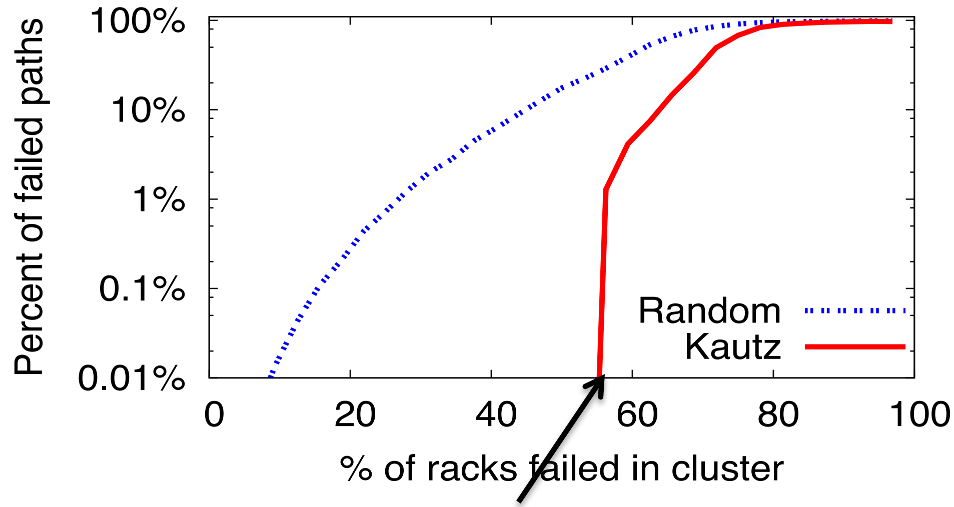
Failure Recovery Results

Random link failures



0.1% paths fail when
20% of links fail

Collocated rack failures



100% connectivity until
>50% collocated racks fail
(leverage far-away siblings)

- Structural fault recovery → good robustness
- Deterministic algorithms → no extra coordinator

Outline

- Motivation
- System design
- Evaluation
  - Testbed
  - Simulation
- Conclusion

Testbed Validation

- Two testbeds
  - HXI: horn antennas
  - Wilocity: 2x8 arrays, affordable for multi-hop

HXI testbed
Horn antenna

Wilocity testbed
2x8 array

- Single link performance
  - Measured per-second TCP throughput over one month
  - Average ~900Mbps (capped by 1Gbps NIC)
  - Standard variation <1% average throughput → as stable as a wired link

Testbed Validation (Multi-hop)

- Without interference
  - Throughput not affected
  - Latency scales with hops

Path Length	TCP Thpt[1]	10KB Latency
2 hops	662Mbps	2.5ms
3 hops	654Mbps	3.1ms
4 hops	665Mbps	3.5ms

Multi-hop path performance

Multi-hop paths have low interference → high throughput and predictable latency.

  - Different channels → no throughput loss
    - 802.11ad defines 3 channels → low self-interference

- Cross-path interference mitigated by node naming

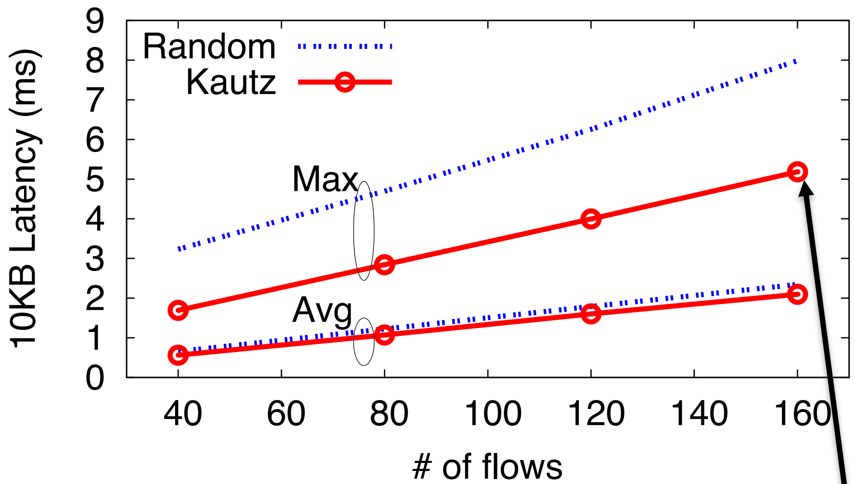
[1]Throughput lower than single link due to software port forwarding overhead

Large-scale Simulation

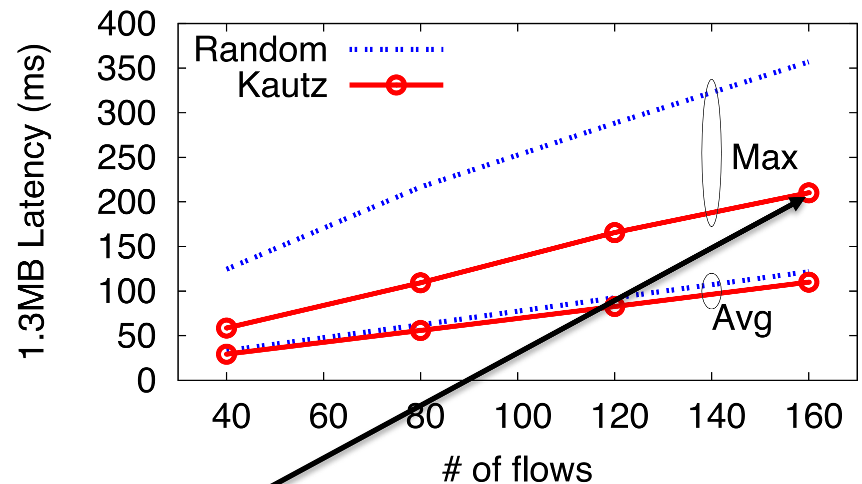
- We implement Angora in NS-3
  - Antenna: horns and arrays
  - 3D beamforming signal reflection
  - 802.11ad PHY/MAC
  - Kautz overlay routing
  - Medium size (320~480 racks) DCN layouts
- Micro-benchmark: path hop count, concurrency, fault-tolerance
- End-to-end performance: single flow, Poisson flows, synchronized flows

End-to-end Performance

- Worst case: synchronized flows



Small message (10KB)



Large messages (1.3MB)

- Tail delay satisfies facilities network requirements
- Structural (Kautz) >> random at tails

Conclusion

- Motivation: build an orthogonal facilities network as a core tool for managing DCN.
- We propose Angora, a Kautz overlay built on 60GHz 3D beamforming links.
- Addressed challenges
  - Wireless interference
  - Robustness to failures
  - Incomplete Kautz graph

Thank you!
Questions?