Controller Envoys: Enabling Scalable Virtualization in DC and HPC Networks

Zhizhen Zhong

Tsinghua University & UC Davis

zhongzz14@mails.tsinghua.edu.cn , zzzhong@ucdavis.edu

23 Feb. 2018

Networks Lab Group Meeting

清華大學
Tsinghua University

UCDAVIS
UNIVERSITY OF CALIFORNIA

Outline

- Network evolution to large-scale instances (DC, HPC)
- Virtualization requirements for large-scale DC and HPC
- Contradiction: large scale vs. network virtualization
- Controller envoys for improving scalability
- Conclusion and discussions

Outline

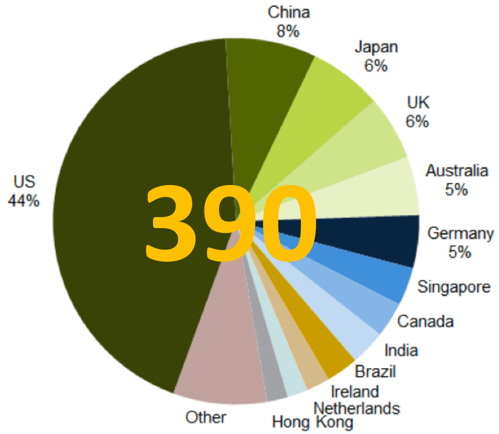
- **Network evolution to large-scale instances (DC, HPC)**
- Virtualization requirements for large-scale DC and HPC
- Contradiction: large scale vs. network virtualization
- Controller envoys for improving scalability
- Conclusion and discussions

DC and HPC: embrace the age of data

- Data Center (DC) is evolving to be hyper-scale (at least 5000 servers and 10,000 square feet of available space), owned by Facebook, Google, Amazon, etc.

Hyperscale Data Center Operators

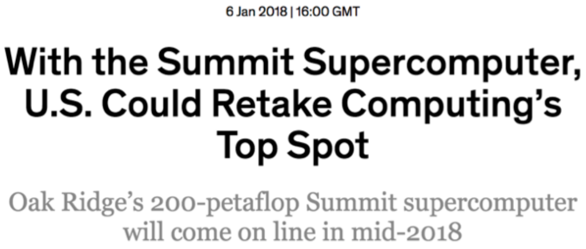
Data Center Locations by Country - December 2017



Source: Synergy Research Group

- Data Center (DC)

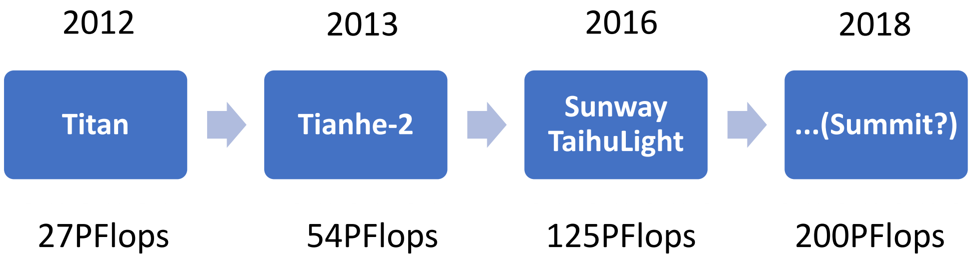
- High-Performance Computing (HPC) is paving to exascale and exaFlops.



6 Jan 2018 | 16:00 GMT

With the Summit Supercomputer, U.S. Could Retake Computing's Top Spot

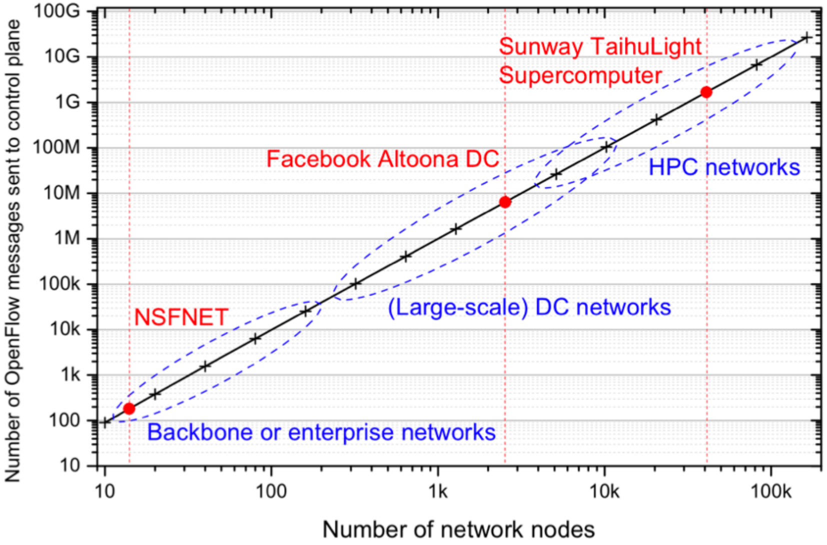
Oak Ridge's 200-petaflop Summit supercomputer will come on line in mid-2018



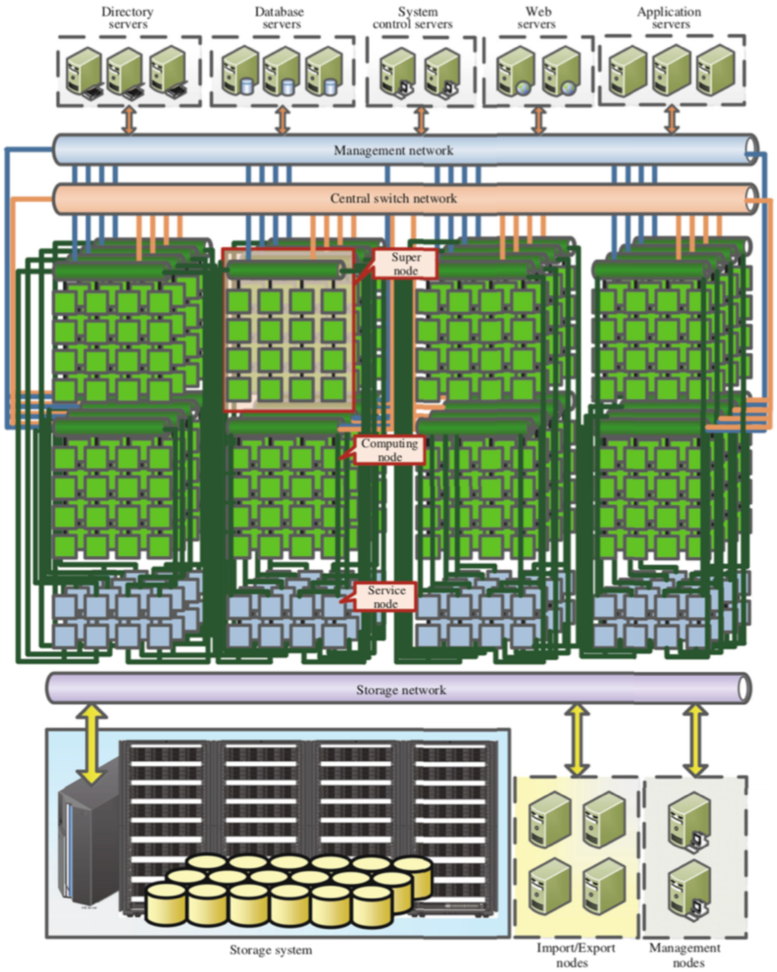
- High-Performance Computing (HPC)

DC and HPC is scaling networks out!

- Conventional optical networks for **Telecom** span only tens or hundreds of nodes.
- NG optical networks for **Datacom**: the size of network is scaling out.
  - Facebook’s new data center in Altoona spans over **2500** network nodes.
  - The top#1 supercomputer, Sunway TaihuLight, spans **40960** networked nodes.


Number of OpenFlow messages sent to control plane
100G
10G
1G
100M
10M
1M
100k
10k
1k
100
10
Number of network nodes
10
100
1k
10k
100k
Sunway TaihuLight Supercomputer
Facebook Altoona DC
HPC networks
NSFNET
(Large-scale) DC networks
Backbone or enterprise networks

Dive into HPC interconnect networks



Networking Facts of Sunway TaihuLight

Network link: 16Gb/s
Super node: 256 Sunway processors
Cabinet: 4 super nodes
Entire systems: 40 cabinets

Outline

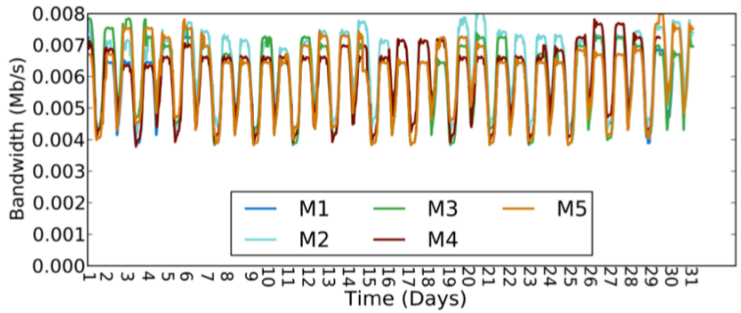
- Network evolution to large-scale instances (DC, HPC)
- **Virtualization requirements for large-scale DC and HPC**
- Contradiction: large scale vs. network virtualization
- Controller envoys for improving scalability
- Conclusion and discussions

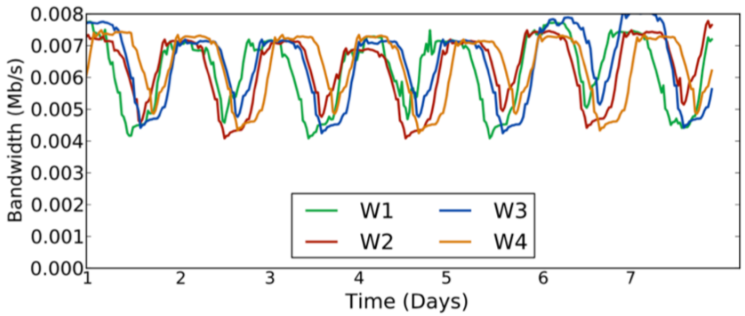
Traffic pattern in large-scale DC

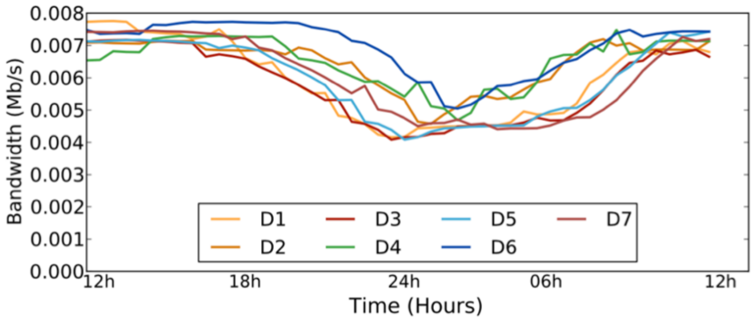
- **Prevalent views on DC traffic characteristics**
- Heavily rack-local distribution
  - A majority of traffic originated by servers (80%) stays within the rack [1,2].
- Bursty and unstable across various timescales
  - traffic is unpredictable at timescales of 150 seconds and longer, it can be relatively stable on the timescale of a few seconds [3].
  - traffic locality varies on a day-to-day basis, it re- mains consistent at the scale of months [4]
- Bimodal pack sizes
  - Packet level statistics, not for us to discuss here

[1] T. Benson, et al. Network traffic characteristics of data centers in the wild.. ACM IMC, 2010.
[2] S. Kandula, et al. . The nature of data center traffic: Measurements & analysiss. ACM IMC, Nov. 2009.
[3] T. Benson, et al. "MicroTE: Fine grained traffic engineering for data centers." Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies. ACM, 2011.
[4] C. Delimitrou, et al. "ECHO: Recreating network traffic maps for datacenters with tens of thousands of servers." Workload Characterization (IISWC), 2012.

Traffic pattern in large-scale DC





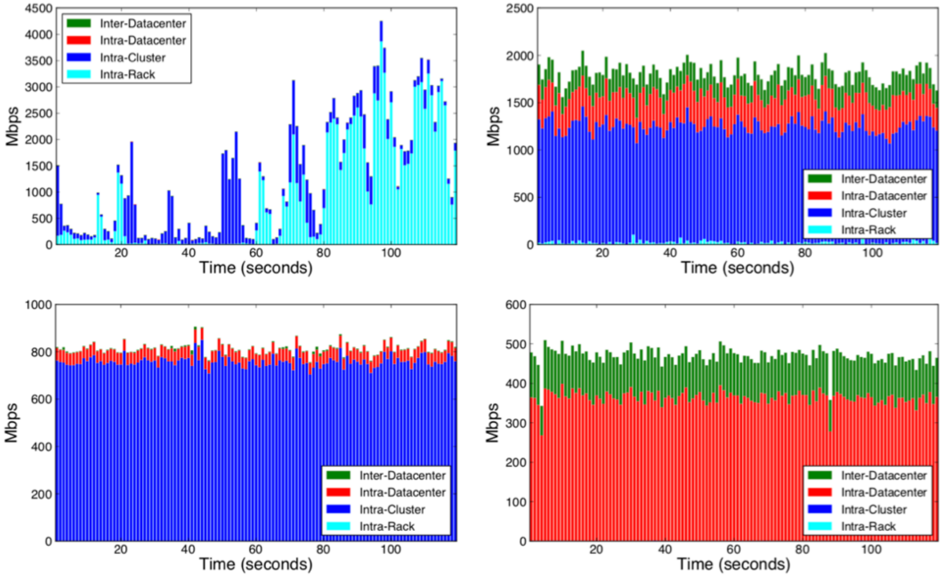


- ➢ DC traffic load variations across months, weeks and days [4].
- ➢ General routine can be found.
  - ➢ Every day rise and fall
- ➢ To better manage DC resource, virtualization technology is crucial.

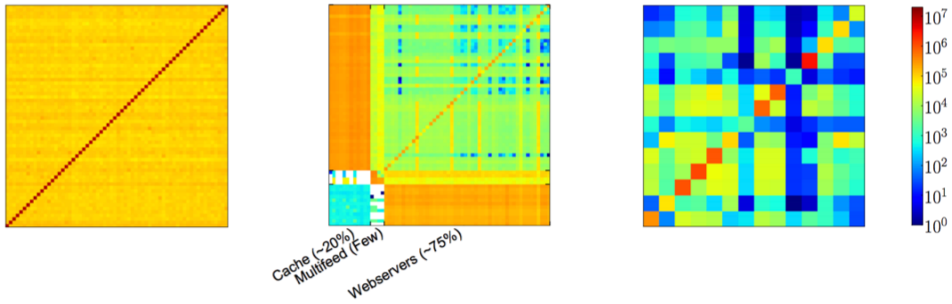
[4] C. Delimitrou, et al. "ECHO: Recreating network traffic maps for datacenters with tens of thousands of servers." Workload Characterization (IISWC), 2012.

Traffic pattern in large-scale DC

- Facebook’ data center traffic pattern is different [5].
  - Most of traffic is inter-rack or inter-cluster.
  - In short-term (hundreds of seconds), traffic load are stable.



- The analysis above on traffic pattern inspires us to design dedicated virtual network to better serve the traffic.

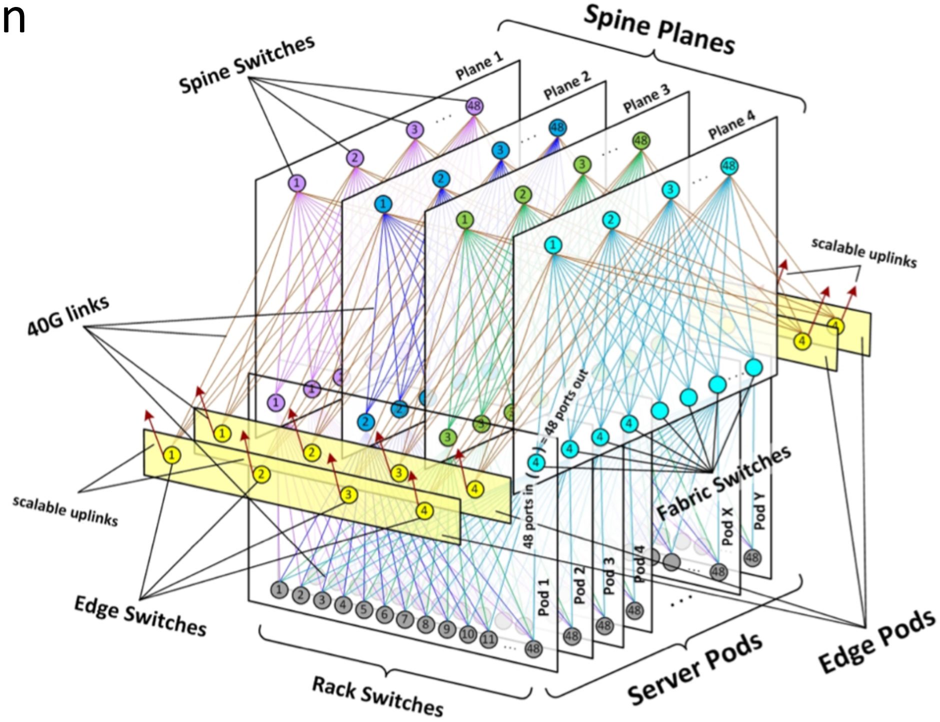


(a) Rack-to-rack, Hadoop cluster (b) Rack-to-rack, Frontend cluster (c) Cluster-to-cluster

[5] A. Roy, et al. "Inside the social network's (datacenter) network," ACM SIGCOMM Computer Communication Review. Vol. 45. No. 4. ACM, 2015.

Virtualization needs for large-scale DC and HPC

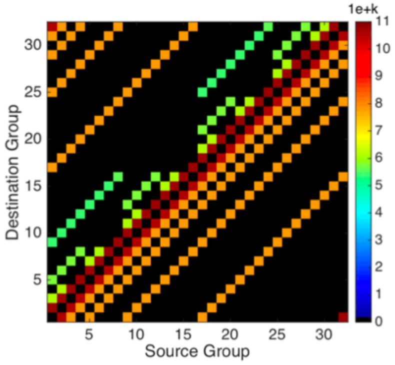
- The close dependency between dedicated resources for specific functions restricts sustainable expansion of traditional IT architecture.
- In intra-DC networks, traffic pattern has daily-fluctuated pattern.
- **Advantages of DC virtualization:**
  - Performance isolation
  - Increased security
  - Application deployable
  - flexible management
  - Support network innovations



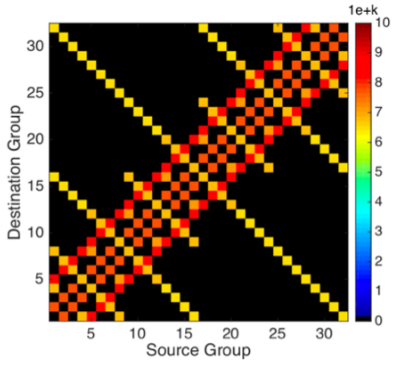
[6] M.F. Bari *et al.*, "Data Center Network Virtualization: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 909-928, 2013.

Traffic pattern in large-scale HPC

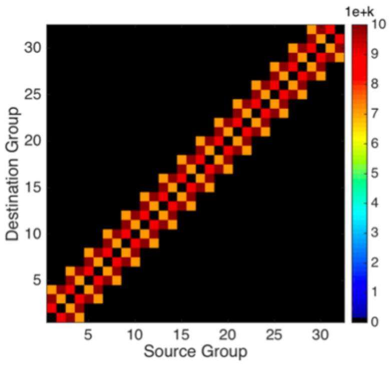
- In 2017, Sunway TaihuLight has finished **2 million** computing tasks. The computing infrastructure is fully multiplexed by time.
- HPC computing tasks have specific communication demands. Each task is calculated during a certain amount of time.
- Node coordination can be satisfied via virtual networks.
- Network virtualization can help increase resource utilization efficiency.



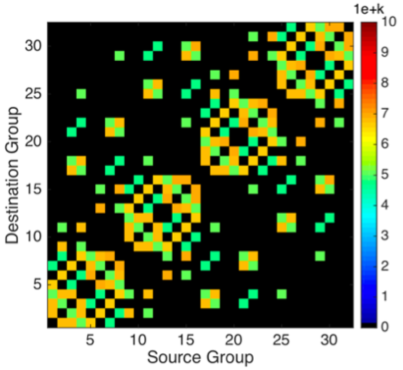
GTC



Nekbone



LULESH



MiniFE

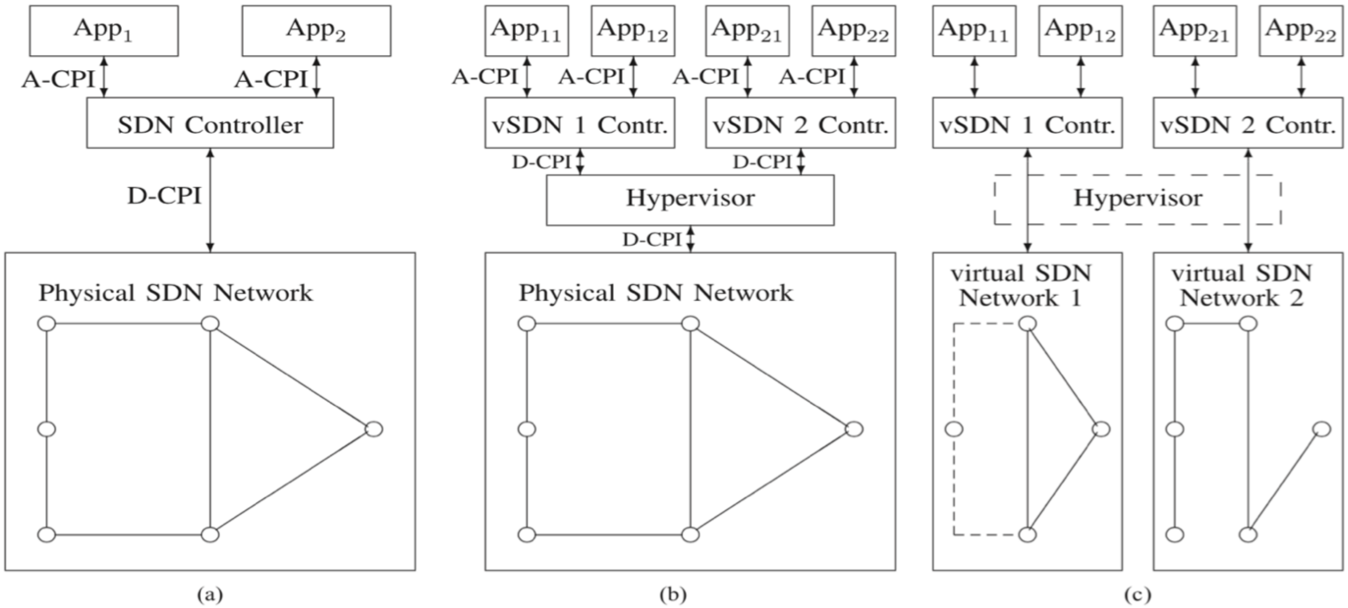
[7] K. Wen, et al. "Flexfly: Enabling a reconfigurable dragonfly through silicon photonics." *High Performance Computing, Networking, Storage and Analysis, SC16: International Conference for*. IEEE, 2016.

Outline

- Network evolution to large-scale instances (DC, HPC)
- Virtualization requirements for large-scale DC and HPC
- **Contradiction: large scale vs. network virtualization**
- Controller envoys for improving scalability
- Conclusion and discussions

Cost of virtualization

- **Bottleneck: virtualization hypervisors**
  - The control signal from all virtual controllers has to go through the only hypervisor!
  - The cost of virtualization: control latency overhead.



Traditional un-virtualized SDN

Virtualized SDN (hypervisor view and vController view)

[8] A. Blenk, et al. "Survey on network virtualization hypervisors for software defined networking." IEEE Communications Surveys & Tutorials 18.1 (2016): 655-685.

[9] A. Blenk, et al. "Control plane latency with SDN network hypervisors: The cost of virtualization." IEEE Transactions on Network and Service Management 13.3 (2016): 366-380.

Outline

- Network evolution to large-scale instances (DC, HPC)
- Virtualization requirements for large-scale DC and HPC
- Contradiction: large scale vs. network virtualization
- **Controller envoys for improving scalability**
- Conclusion and discussions

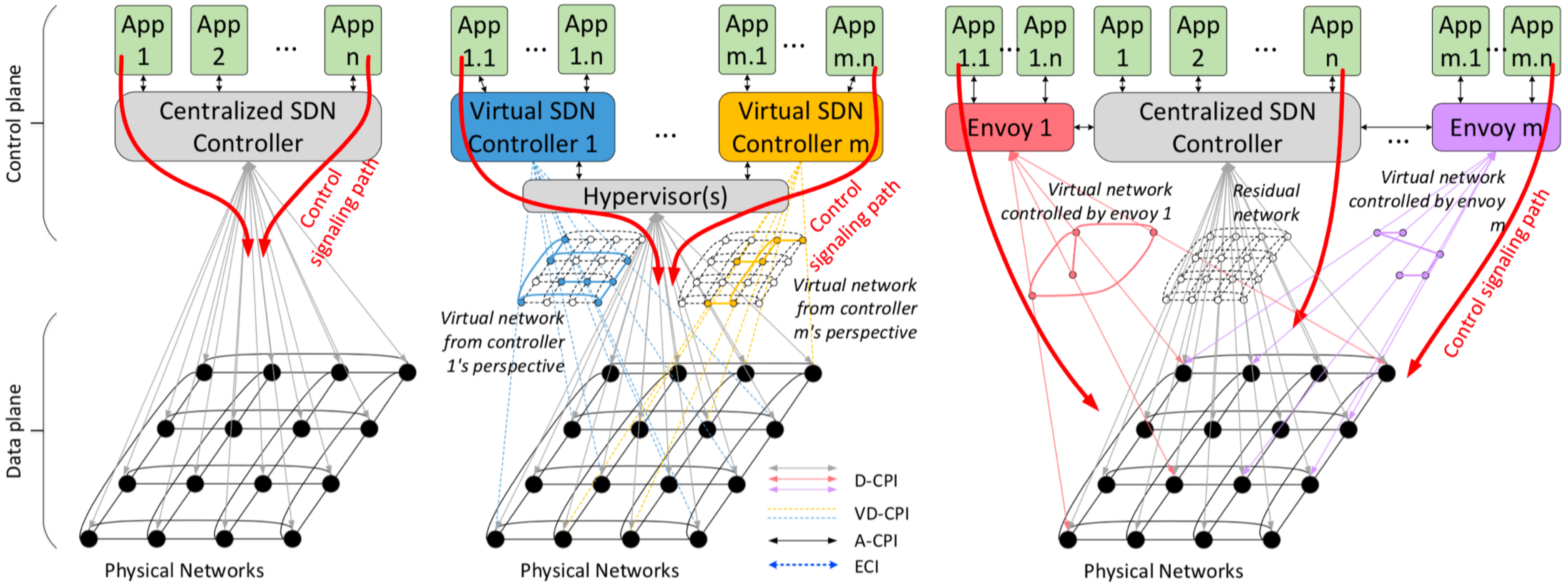
Revisiting network virtualization

- Managing virtualization is like managing a company:

Customer	**Network application**
Board of the company	**Controller**
Factory and staff	**Network infrastructure**

- When this company is becoming really huge? Does every matters about hundreds of factories and millions of staffs need to be decided by a central board?
- Example: UC Davis wants Cisco to renovate its campus networks.
- UC Davis do not care which group of people in Cisco to do this job.
- Cisco board do not need to discuss the details of this project all the time.
- **Cisco board: let do it! Then send a envoy to take full change of it.**

Virtualization architecture evolution

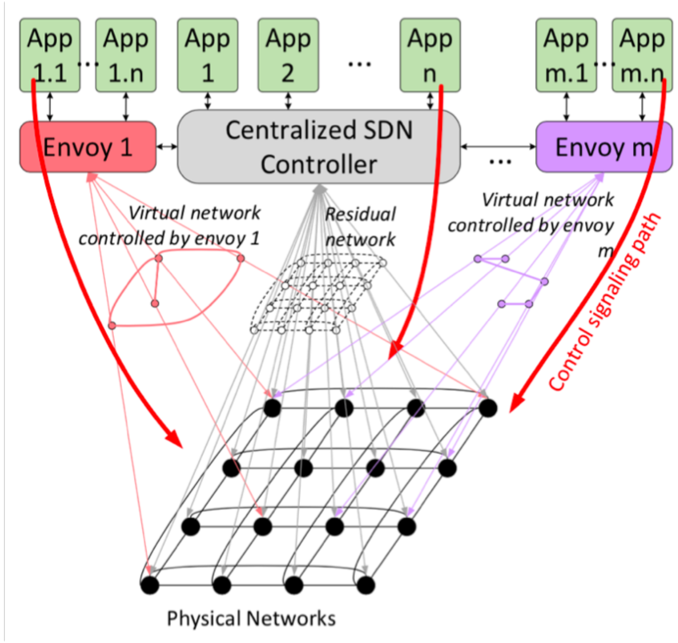


(a) Conventional SDN.

(b) SDN virtualization with hypervisors.

(c) Scalable SDN virtualization with controller envoys.

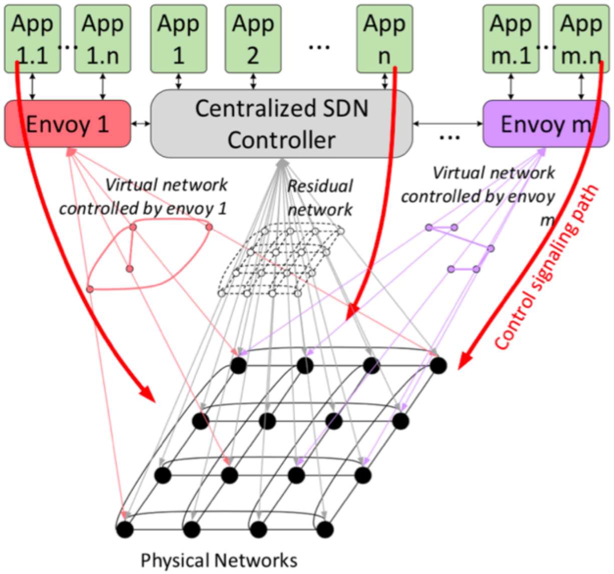
Key Features of Controller Envoys



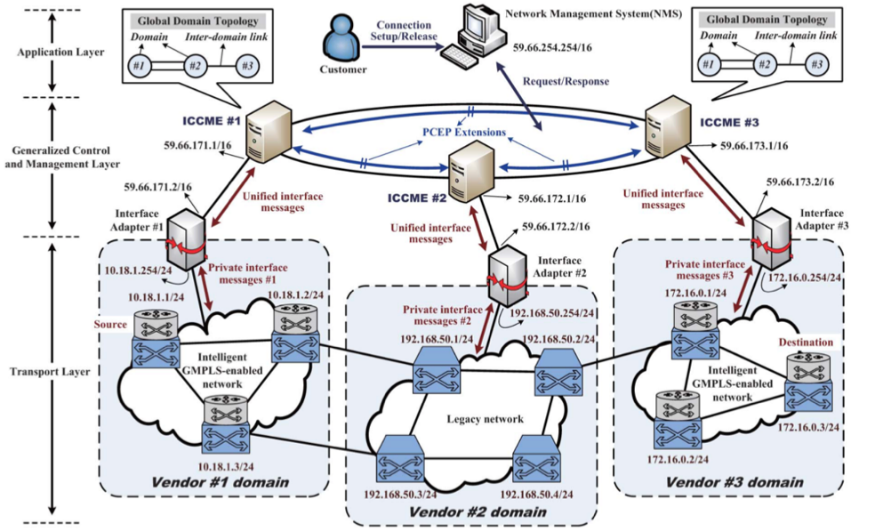
- **Physically-distribute located:**
  - A envoy can choose its locate best fit for VN.
- **Per-task assignment:**
  - for a specific DC virtual network or HPC computing task, assign a envoy to take charge of the VN. So, it is a dynamic process.
- **Intra-VN autonomy:**
  - Envoy has all the information regarding the VN when assigned. The envoy can fully control the inside traffic flows of VN, without bothering the centralized controller. The controller will view the VN as a node.
- **Inter-VN peering:**
  - For inter-VN traffic, envoys can peering with each other and the controller.

More elaborations: Envoy vs. Domain Controller

Controller Envoys for large-scale networks



Hierarchical controllers for large-scale multi-domain networks

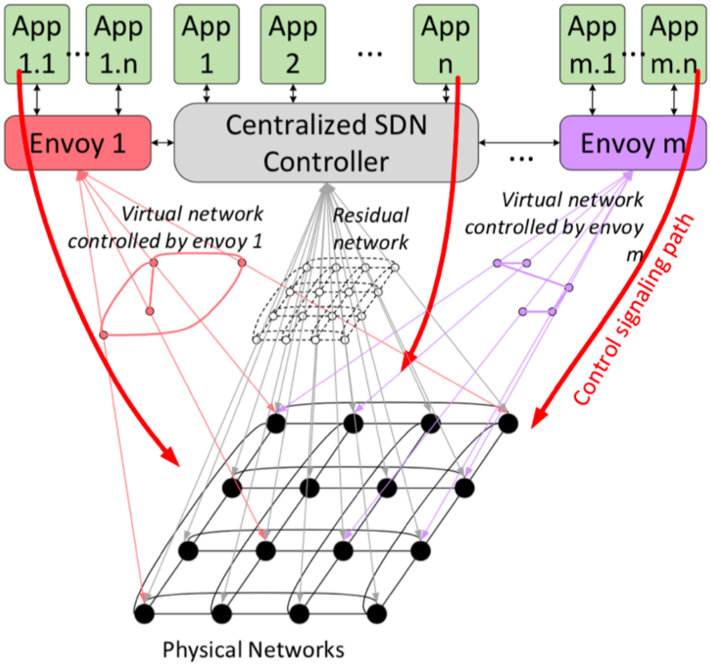


- **Similarities**: Envoys acts like a virtual domain controller. We dynamically allocate virtualized multi-domain-like networks inside a single-domain network, to alleviate the central controller’s load.
- **Difference**: domain networks, and domain controller are fixed.

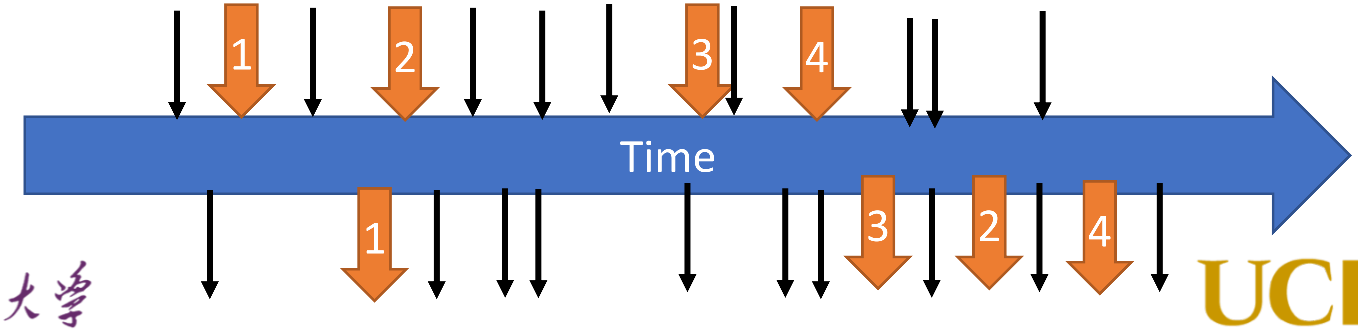
Outline

- Network evolution to large-scale instances (DC, HPC)
- Virtualization requirements for large-scale DC and HPC
- Contradiction: large scale vs. network virtualization
- Controller envoys for improving scalability
- **Discussions and research ideas**

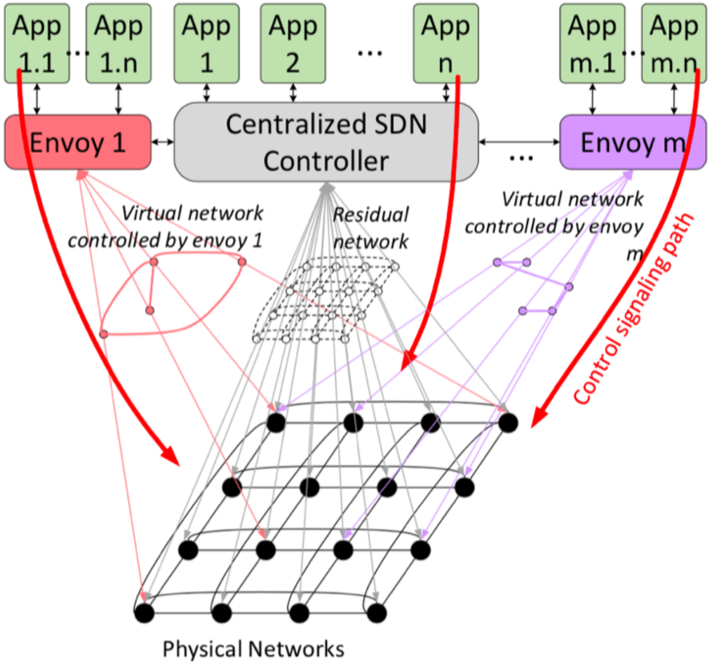
Research problems



- Assumption: hybrid request model (realistic)
  - VN requests: sparse arrival
  - Traffic requests:
    - Intra-VN: high intensity
    - Inter-VN: low intensity
- Given VN requests and its mapping, evaluate how envoys can improve intra-VN and inter-VN traffic requests control latency.
- Envoys acts like a virtual domain controller.



Research problems



- **Static ILP formulation**: When VN request arrives, we optimize the location of its envoy to minimize control latency
  - This is a static problem, because traffic pattern of DC and HPC can be acquired in advanced, so we can schedule in advance.
- **Dynamic event-driven simulation**: we analyze how the envoy control the running traffic inside or through the VN after it is assigned.
  - This is a dynamic problem, because traffic requests com and go in a relative dynamic manner, and cannot be predicted.

Expected contributions

- **The introduction of the novel envoy scheme for network virtualization at large.**
  - A novel architecture for large scale network virtualization.
- **Design protocol for envoy peering with each other and the controller**
  - Enabling technology.
- **Formulate the problem of envoy location selection given VN and its mapping.**
  - Enabling technology.
- **Test the performance of envoy architecture under dynamic network scenarios**
  - Numerical evaluations, envoy architecture reduce control latency overhead.

Expected conclusion

- In this work, we introduce controller envoys to take charge of specific virtual networks (VN) instead of the central controller during the lifetime of VN.
- The envoy can communicate with the central controller by special-designed protocol.
- Controller envoy is assigned in a per-VN manner, and it can fully control the intra-VN traffic, alleviating the signaling burden of the central controller.
- For inter-VN traffic, the cooperation among the central controller and involved envoys is needed.
- Simulation results prove that controller envoys can significantly reduce controller burden for network virtualization.

Thank you for attention!

Zhizhen Zhong

Tsinghua University & UC Davis

zhongzz14@mails.tsinghua.edu.cn , zzzhong@ucdavis.edu

23 Feb. 2018

Networks Lab Group Meeting