FlexNet: Enabling Flexible Datacenter Networks

Rodrigo Fonseca (Brown University)

With Da Yu, Shuwen Sun, Raja Sambasivan, Luo Mai, Orran Krieger, Piyanai Saowarattitada, Jason Hennessey

Datacenter Networking Innovation

"New" network topologies

CLOS Fabrics

FatTrees

Randomized Topologies

Datacenter Networking Innovation

New Control Planes

SDN
Network Virtualization
NFV

Datacenter Networking Innovation

New Transport Protocols

DCTCP

PDQ, D3, D^2TCP

pFabric, Qjump, NDP

Datacenter Networking Innovation

New Scheduling Algorithms

Coflows

…

Datacenter Networking Innovation

New “crazy” ideas

Jellyfish topology
Disco Ball
Flywheels
…

Datacenter Networking Innovation

Much better than the Internet!

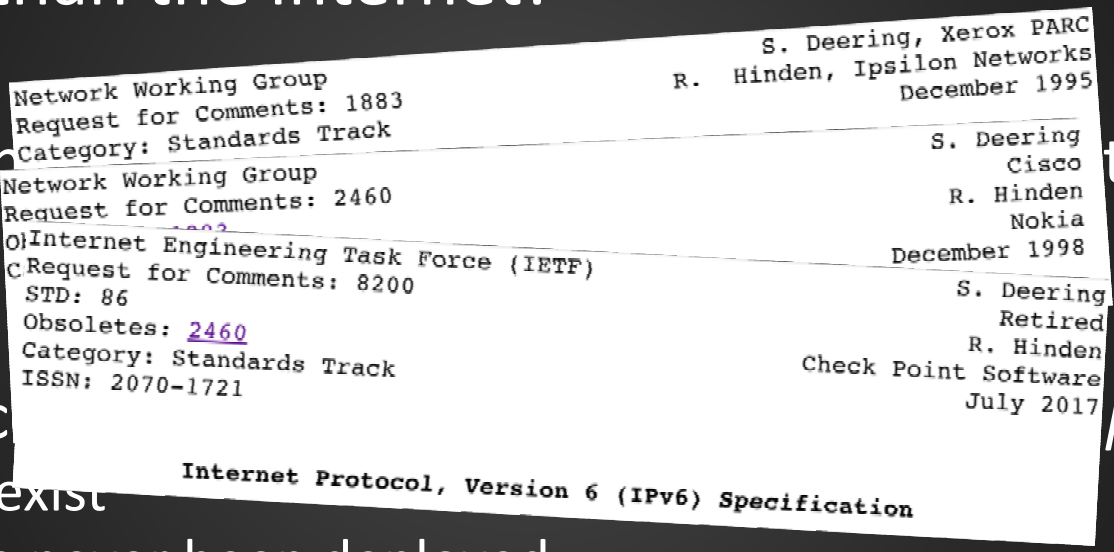
Why?

- Single admin

Really?

- Most protoc *patible*
  - Cannot coexist
  - Many have never been deployed
  - Only “forklift upgrades”: *next* datacenter

A single datacenter is *more* ossified than the Internet!



Network Working Group
Request for Comments: 1883
Category: Standards Track

S. Deering, Xerox PARC
R. Hinden, Ipsilon Networks
December 1995

Network Working Group
Request for Comments: 2460

S. Deering
Cisco
R. Hinden
Nokia
December 1998

Internet Engineering Task Force (IETF)
Request for Comments: 8200
STD: 86
Obsoletes: 2460
Category: Standards Track
ISSN: 2070-1721

S. Deering
Retired
R. Hinden
Check Point Software
July 2017

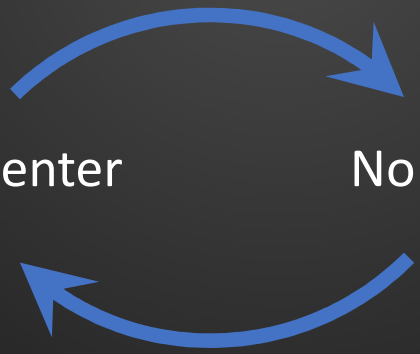
Internet Protocol, Version 6 (IPv6) Specification

Lack of Flexibility

Usually single topology, congestion control, scheduling

Really high bar for new proposals

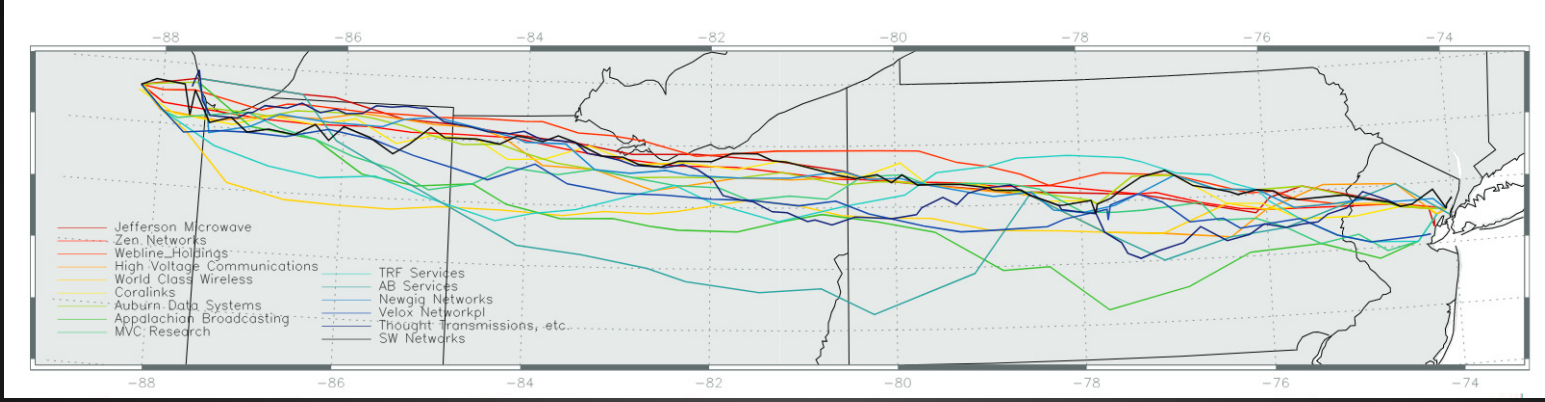
- Must be good for *all* kinds of traffic

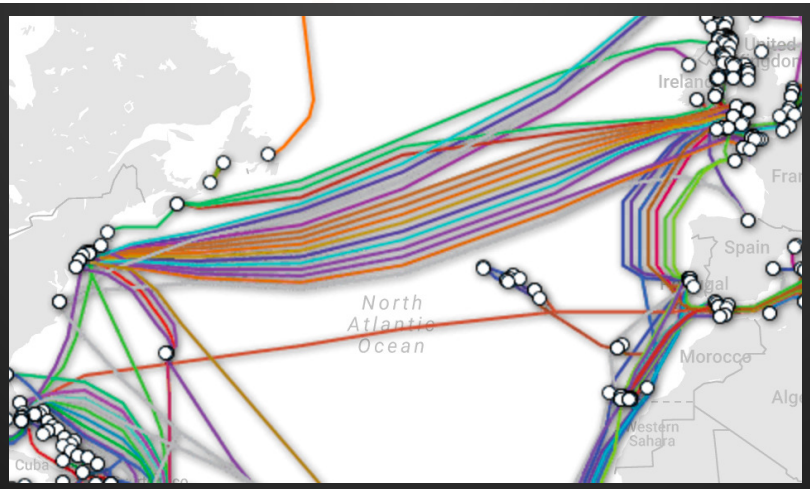


Single protocol across entire datacenter

No standard for application requirements

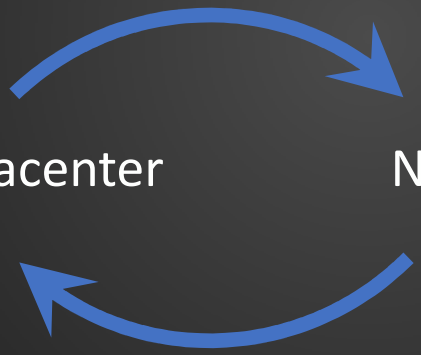
We can learn from the Internet





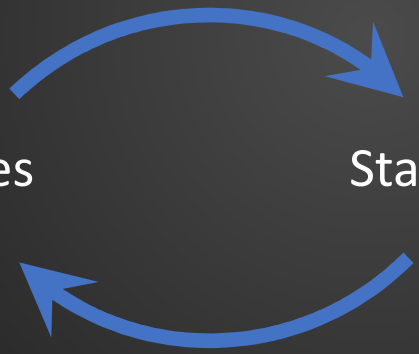
[1] Gregory Laughlin et al., ”Information Transmission Between Financial Markets in Chicago and New York”, arXiv:1302.5966 [q-fin.TR]
[2] https://www.submarinecablemap.com/

Single protocol across entire datacenter



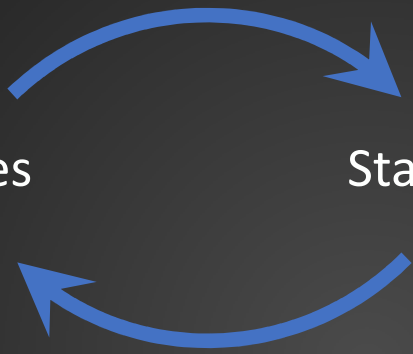
No standard for application requirements

Paths with different properties



Standard way for users to express requirements

FlexNet Datacenter Architecture



Paths with different properties

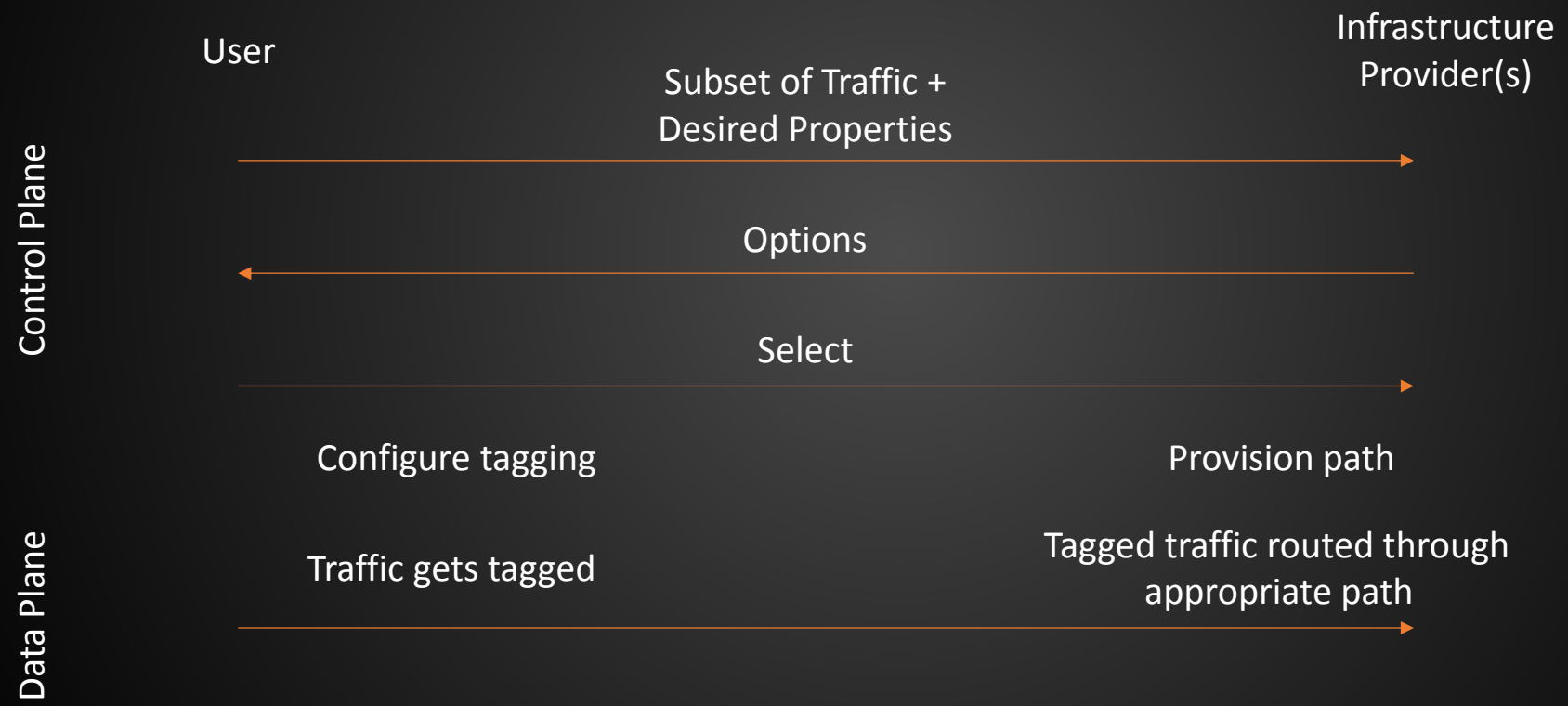
Standard way for users to express requirements

Architecture-assisted *tagging* of packets

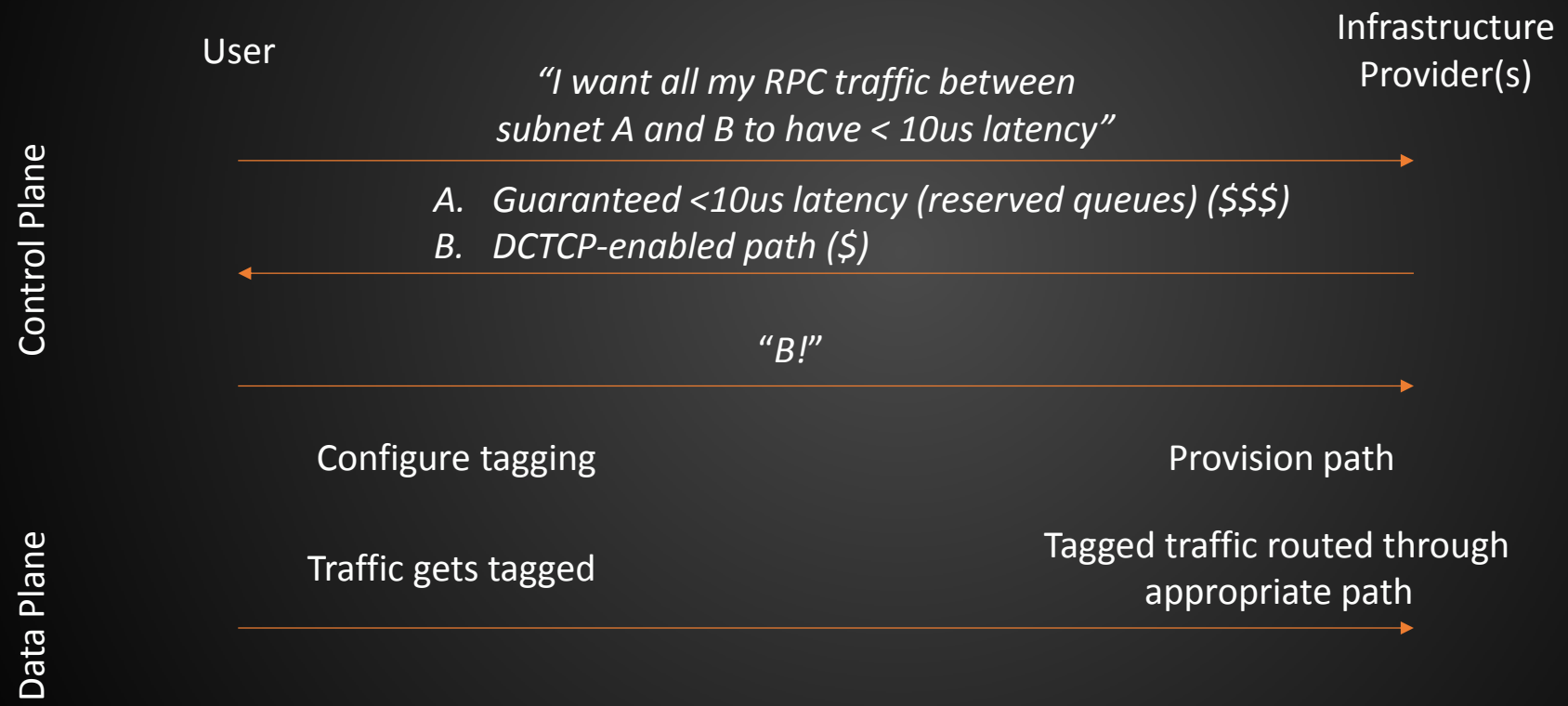
Out-of-band negotiation

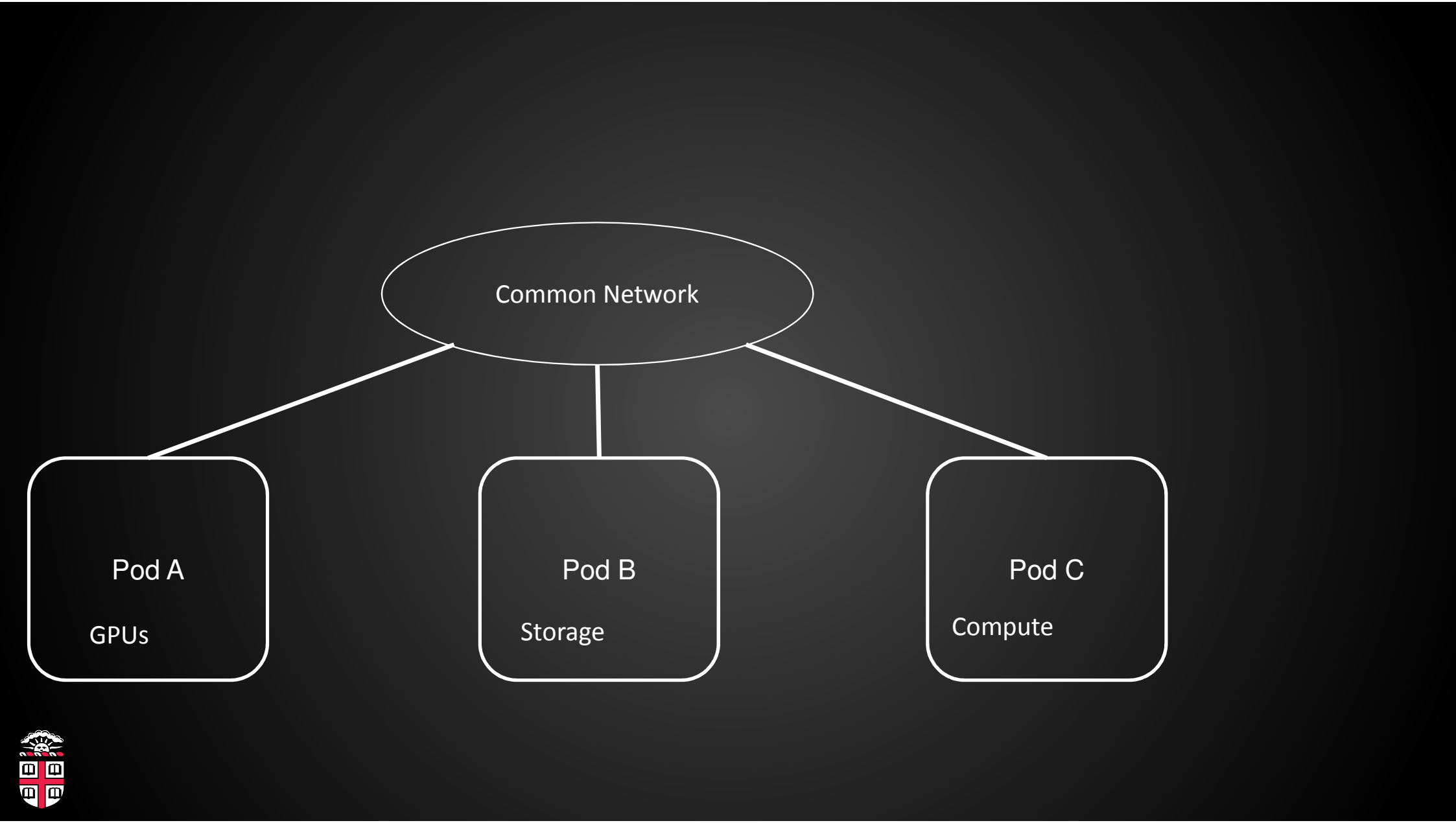
Flexible forwarding

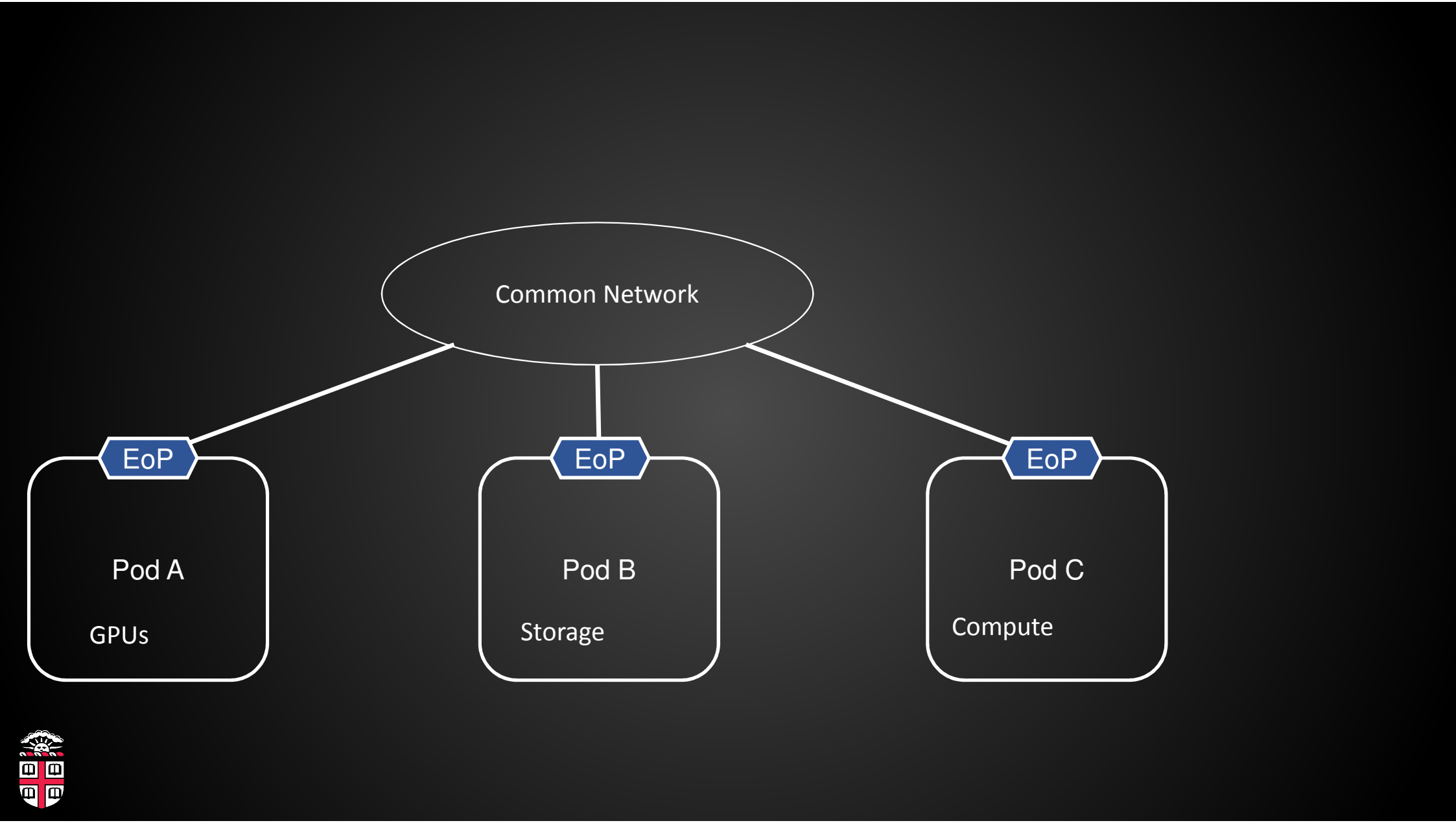
FlexNet Architecture

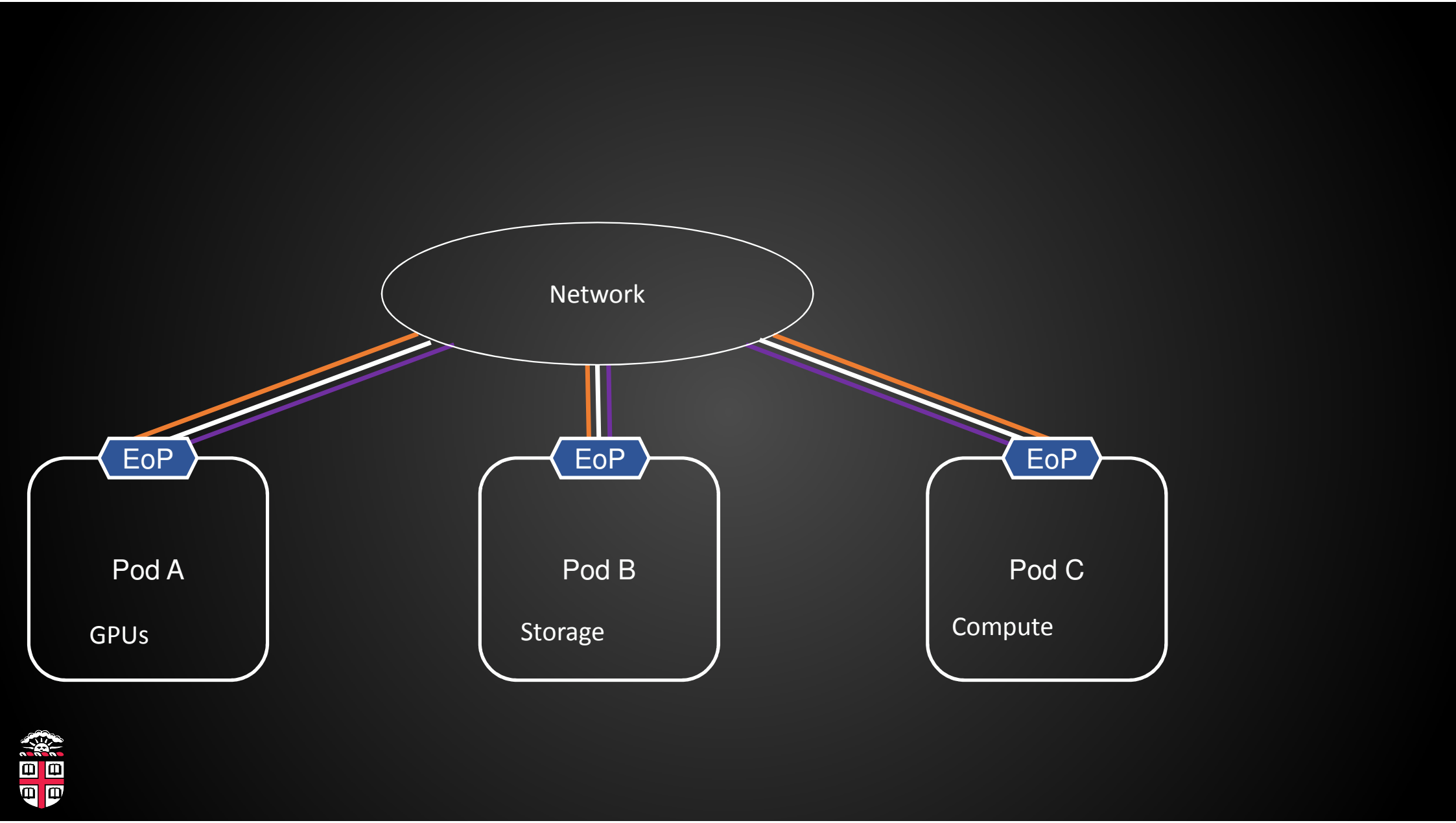


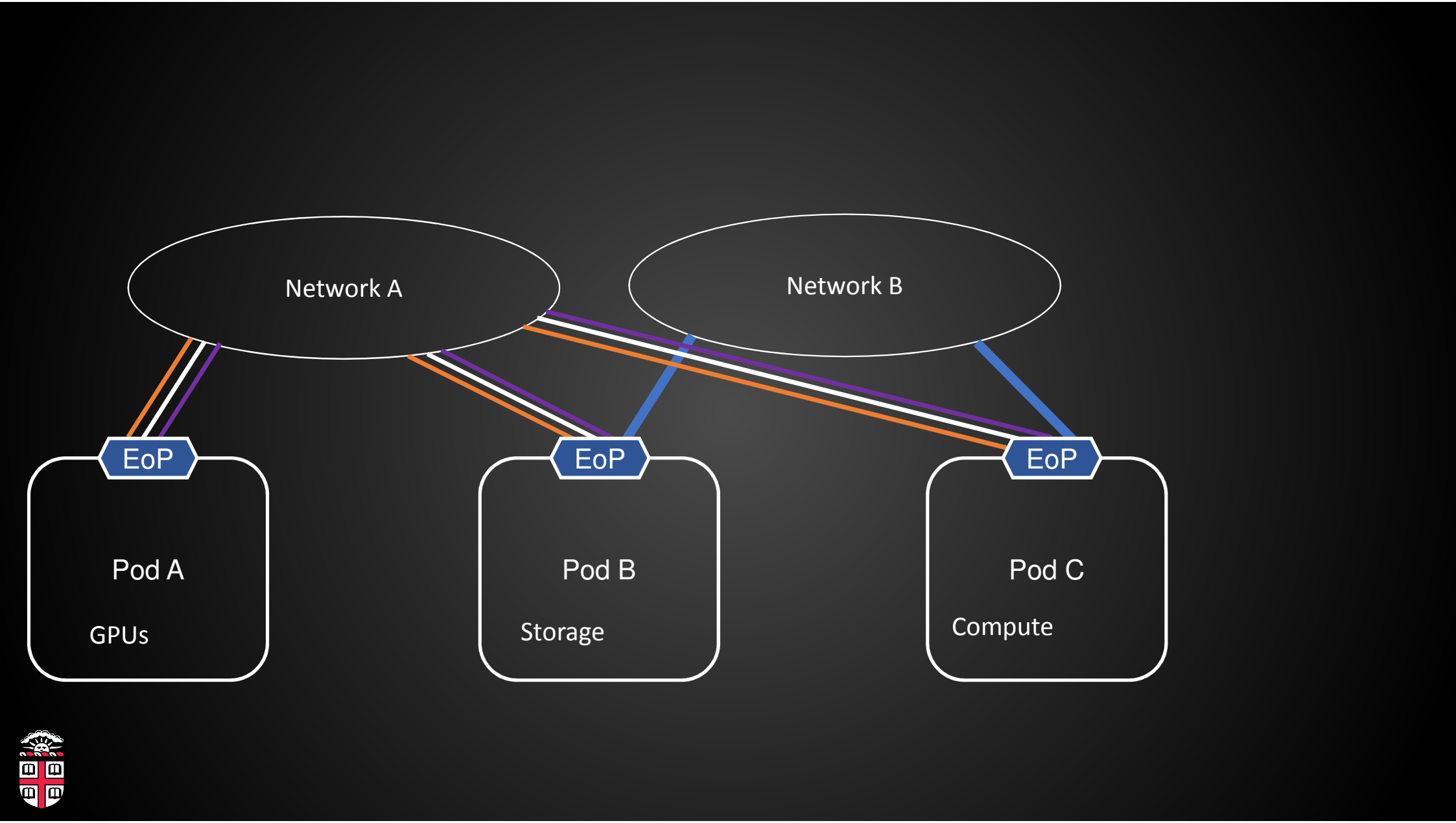
FlexNet Architecture

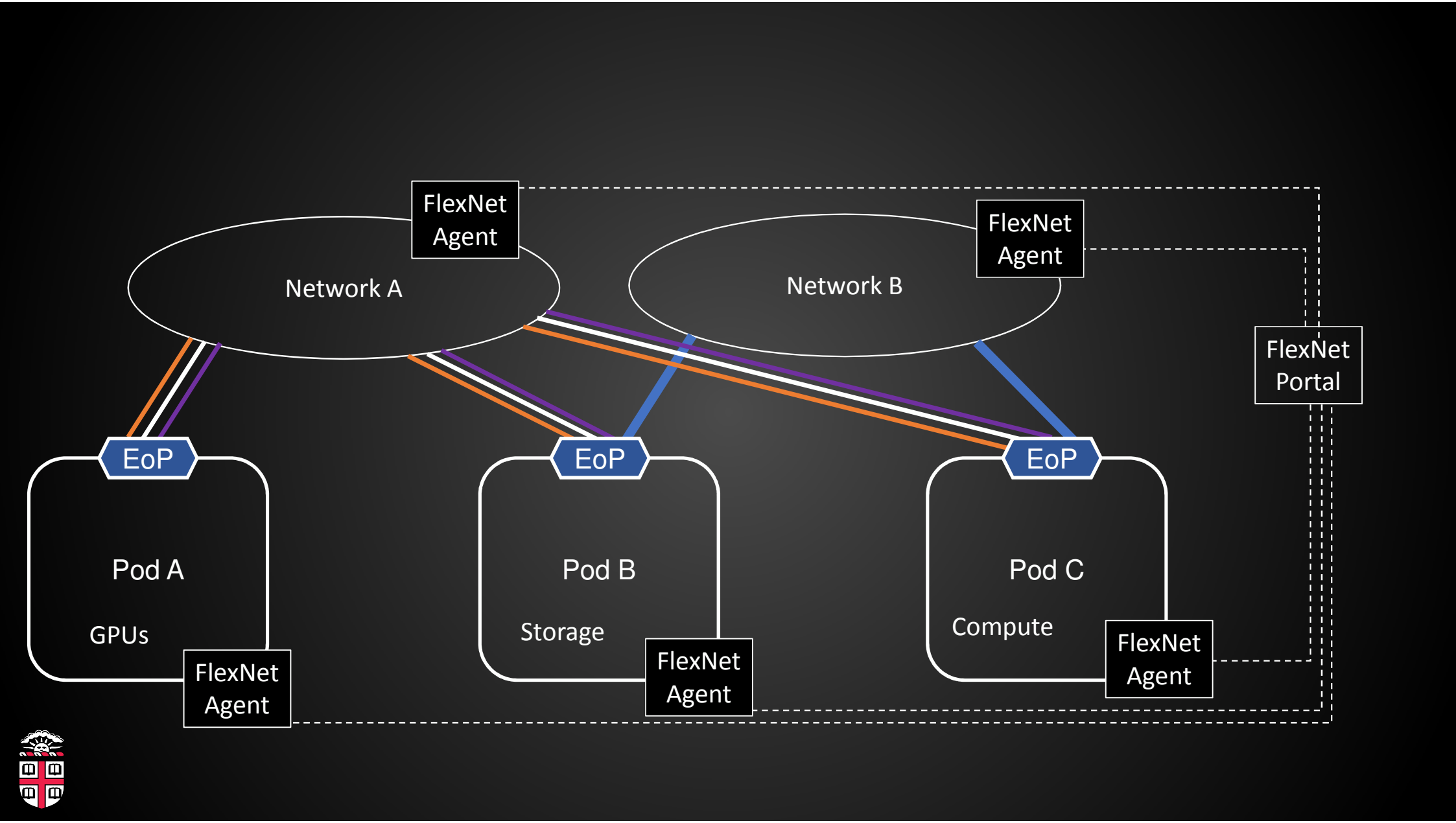



Common Network
Pod A
GPUs
Pod B
Storage
Pod C
Compute


Common Network
EoP
EoP
EoP
Pod A
GPUs
Pod B
Storage
Pod C
Compute


Network
EoP
EoP
EoP
Pod A
GPUs
Pod B
Storage
Pod C
Compute

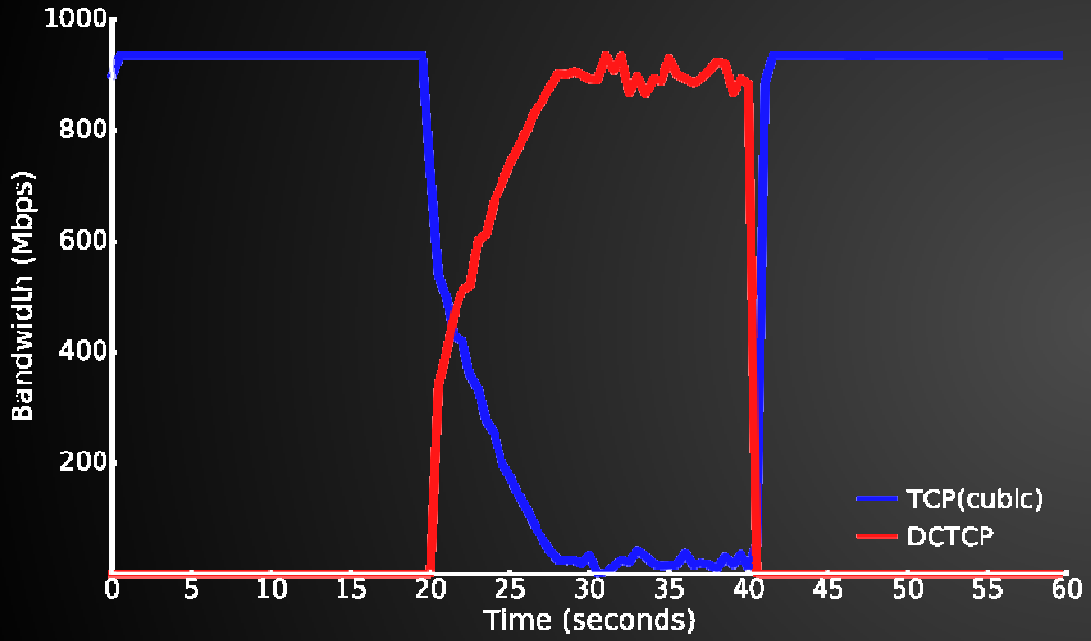

Network A
Network B
EoP
EoP
EoP
Pod A
GPUs
Pod B
Storage
Pod C
Compute

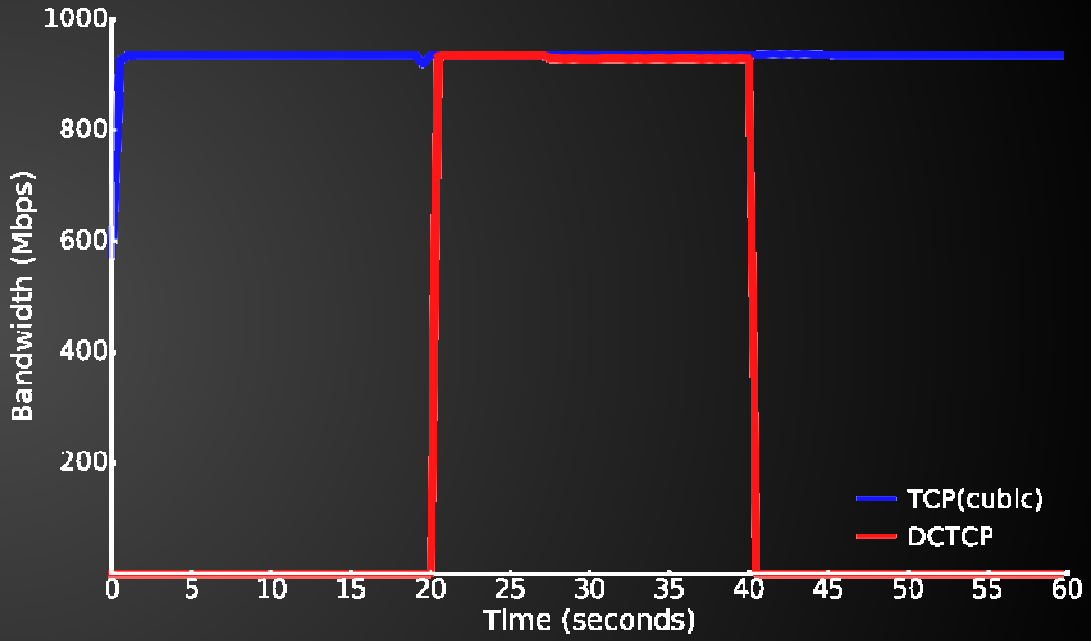

FlexNet Agent
Network A
FlexNet Agent
Network B
FlexNet Portal
EoP
EoP
EoP
Pod A
GPUs
Pod B
Storage
Pod C
Compute
FlexNet Agent
FlexNet Agent
FlexNet Agent

Current Prototype

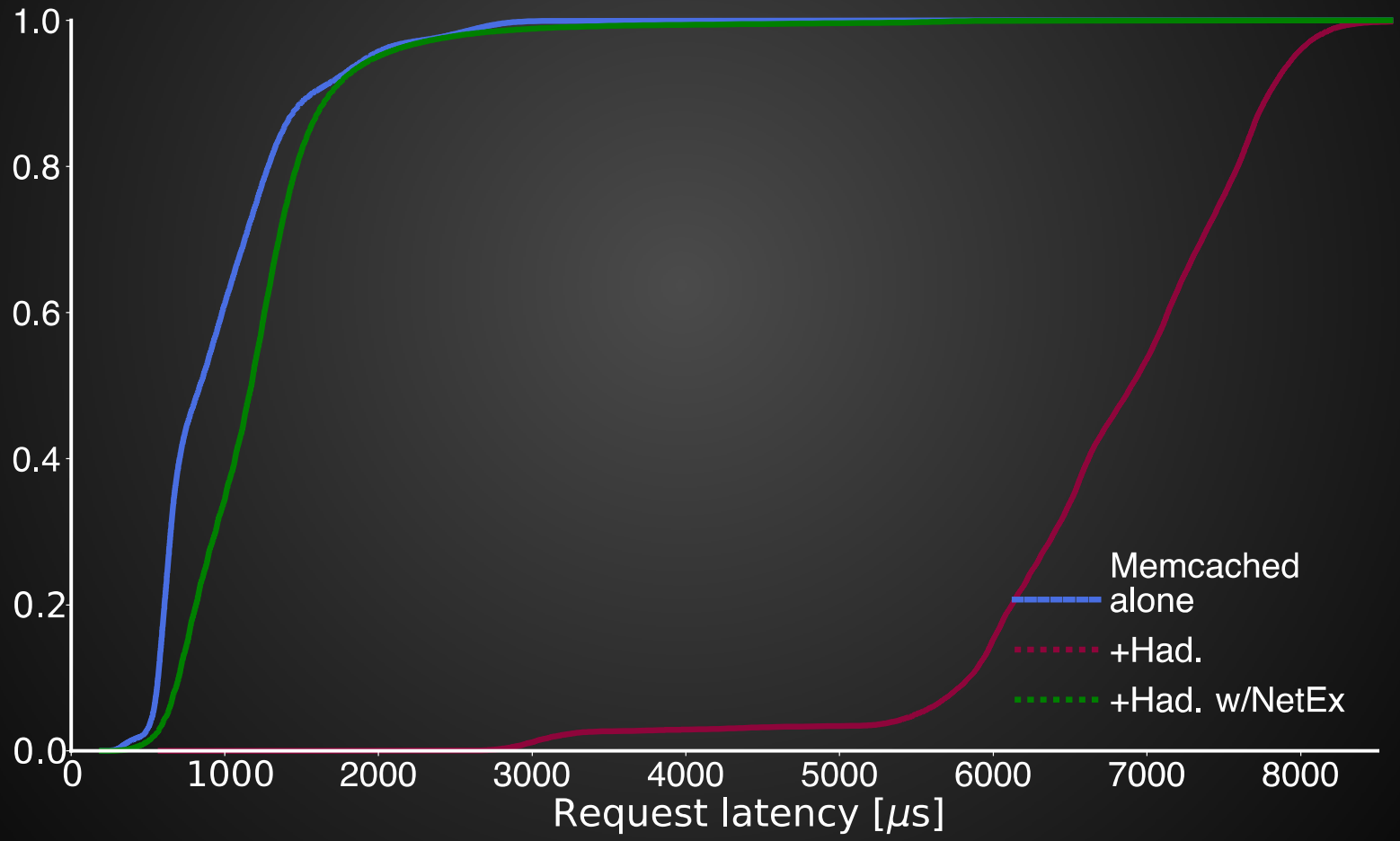
- Complete end-to-end prototype
- Northbound portal in Python to negotiate paths
  - Configures tagging (pod), forwarding (EoP), paths (network)
- Commodity SDN switches + OpenVSwitch
  - VLANs at the pods
  - Nested MPLS to select Network / Pathlet

Evaluation: DCTCP vs Cubic





Memcached + Hadoop



FlexNet

True path diversity + Way to express client requirements

Architecture-assisted Tagging as *Narrow Waist*

Restrictions:

Currently pods all use Ethernet

For some protocols, no bottlenecks in pods

Multiple: paths, networks, providers

Flexibility, innovation, evolution!

Thank you!

https://massopen.cloud/blog/networking/

rfonseca@cs.brown.edu