It's Time for Low Latency

Steve Rumble, Diego Ongaro, Ryan Stutsman,
Mendel Rosenblum, John Ousterhout

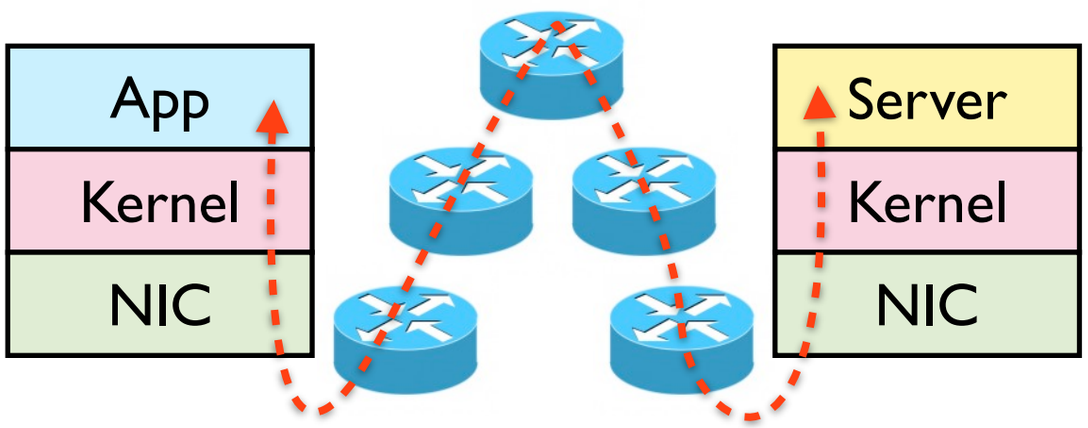
Stanford University

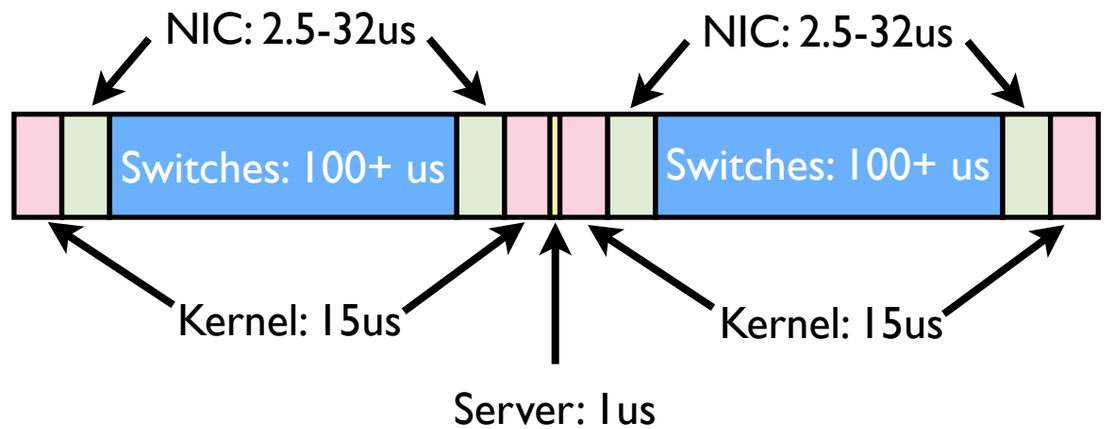
Future Web Applications Need Low Latency

- They will access more bytes of data
  - Bandwidth problem
  - Commodity net bandwidth has increased > 3,000x in 30 years
- But also more pieces of inter-dependent data
  - Latency problem
  - Commodity net latency has decreased only ~30x in 30 years
- Facebook is a glimpse into future applications
  - Huge datasets, DRAM-based storage, small requests, random dependent data accesses, low locality
  - Dependent on network latency:
    Can only afford 100-150 dependent accesses per page request

Datacenter Latency Is Too High

Simple RPCs take 300-500us in current datacenters





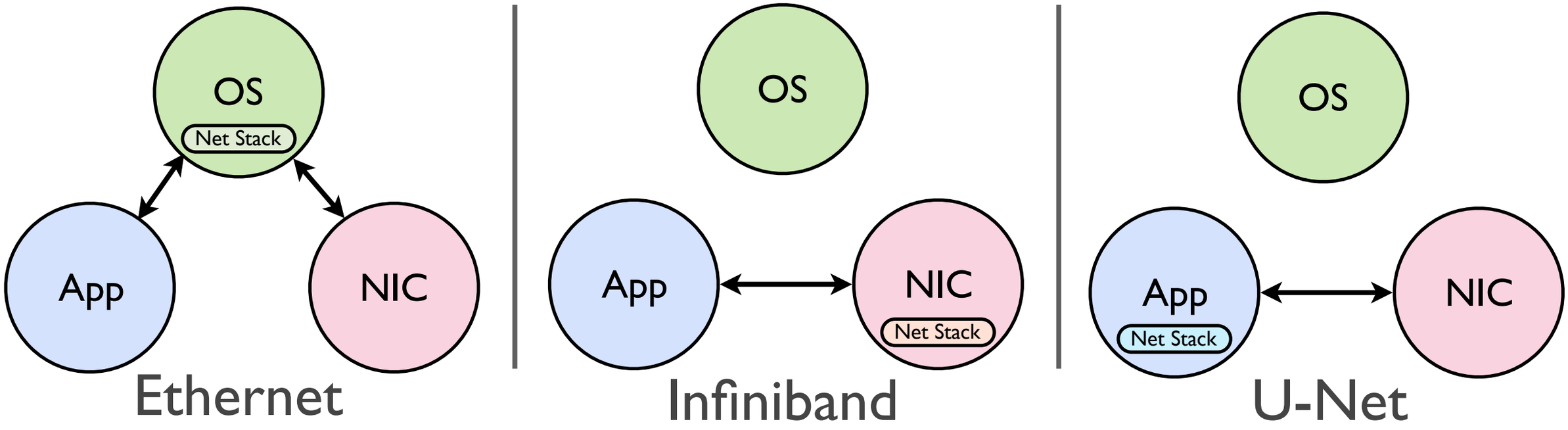
Component	Delay	Round-Trip
Switch	10-30us/hop	100-300us
NIC	2.5-32us	10-128us
OS Net Stack	15us	60us
Server Code	1us	1us
Speed of Light	5ns/m	< 2us

Not limited by server execution or propagation delay!

On The Cusp Of Low Latency

- Low latency available in the HPC space (Infiniband)
  - 100ns switches
  - < 1us NIC latencies
  - OS Bypass (U-Net style)
  - *But*, won't displace Ethernet
- Some migration into commodity Ethernet space
  - Fulcrum Microsystems, Mellanox: Sub-500ns switches
  - RDMA on commodity NICs (e.g. iWarp)
- Now we need to pull in the rest of the ideas
  - Let's get the OS community involved and do it right
  - Goal: 5-10us RTTs in the short term

An Opportunity To Define The Right Structure



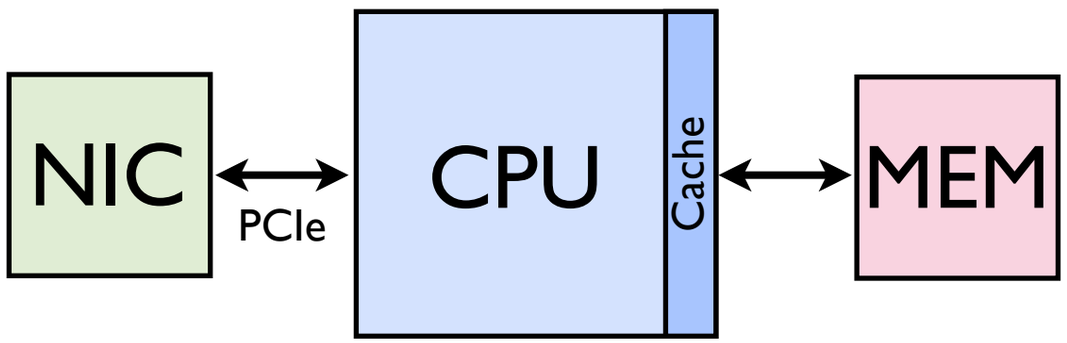
- Re-think APIs: Apps need speed and simplicity
  - Infiniband verbs too complex, RDMA too low-level
  - Developers used to sockets, but can we make them fast?
- Network Protocols
  - Can we live with TCP? (Needs in-order delivery, Slow stacks)
  - How do we scale low-latency to 100,000+ nodes?
  - Closed datacenter ecosystem makes new protocols feasible

Getting The Lowest Possible Latency

The NIC will become the bottleneck under 10us

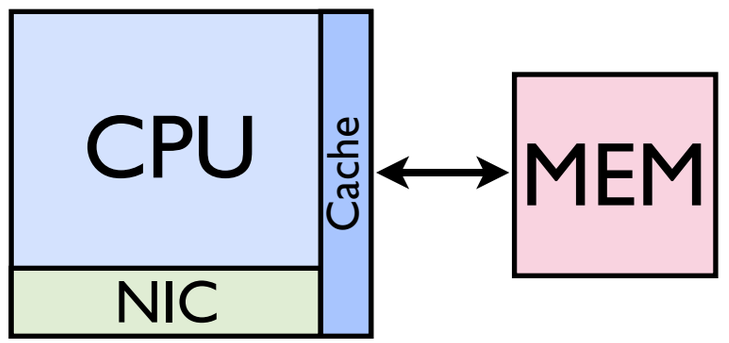
- 500ns round-trip propagation in 50m diameter
- 1us round-trip switching latency (10 x 100ns hops)
- Even fast NICs take nearly 2us on each end!

Today:



PCIe accesses & memory accesses too slow

5-10 Years:



Transmit/Receive directly from/to cache

One microsecond RTTs possible in 5-10 years

Low Latency Is Up To Us

- Low latency is the future of web applications
- If we don't take action to make it happen, we risk:
  - Not getting it at all, or
  - Missing the opportunity to re-architect (and getting something that sucks)