Accelerating Network Function Virtualization and Service Function Chain Processing for Emerging Services & Edge Computing

Zhi-Li Zhang (zhzhang@cs.umn.edu)

Associate Director for Research, Digital Technology Center
Qwest Chair Professor & McKnight Distinguished University Professor
Department of Computer Science and Engineering

University of Minnesota

Collaborative NSF ICE-T Project with Dr. Prof. Arturo Azcorra of UC3M/IMDEA Network (EU H20 5G-EVE/5G-VNNI Projects)

Commercial 5G Is Here!

- Verizon deployed 1st commercial (mmWave) 5G in US in downtown Minneapolis & Chicago in Spring 2019, now rolled out to more cities
- AT&T, T-Mobile and Sprint (sub-GHz) have also deployed their own

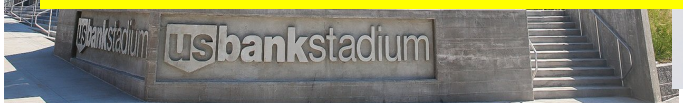
- We have conducted a first measurement study of commercial 5G service deployment in US
  - Verizon, T-Mobile, Sprint in MPLS, CHI, ATL, …

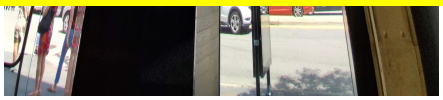
See our paper (earlier version) at http://arxiv.org/abs/1909.07532

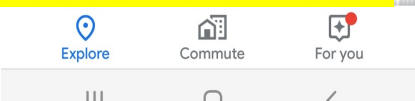
Commercial 5G Is Here!

- Verizon deployed 1st commercial (mmWave) 5G in US in downtown Minneapolis & Chicago in Spring 2019, now rolled out to more cities

- We conducted a first measurement study of commercial 5G service deployment in US
  - Verizon, T-Mobile, Sprint in MPLS, CHI, ATL, …
- Main Take-Aways
  - Yes, 5G has the potential to support exciting new apps (up to 2 Gbps per mobile device, w/ 8 TCP connection)
  - Hugh implications on networking/edge computing: likely pushing bw bottlenecks inside the core; a lot of new challenges!



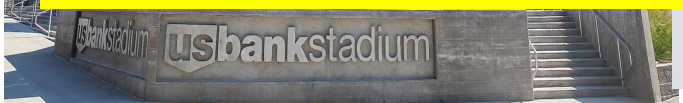


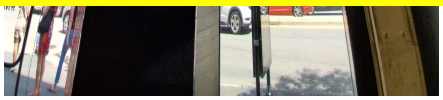
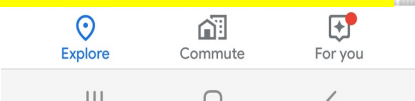


Commercial 5G Is Here! Or Really?

Existing 5G Commercial Deployment is NSA ("non-standalone")
i.e., attached to existing 4G LTE core networks

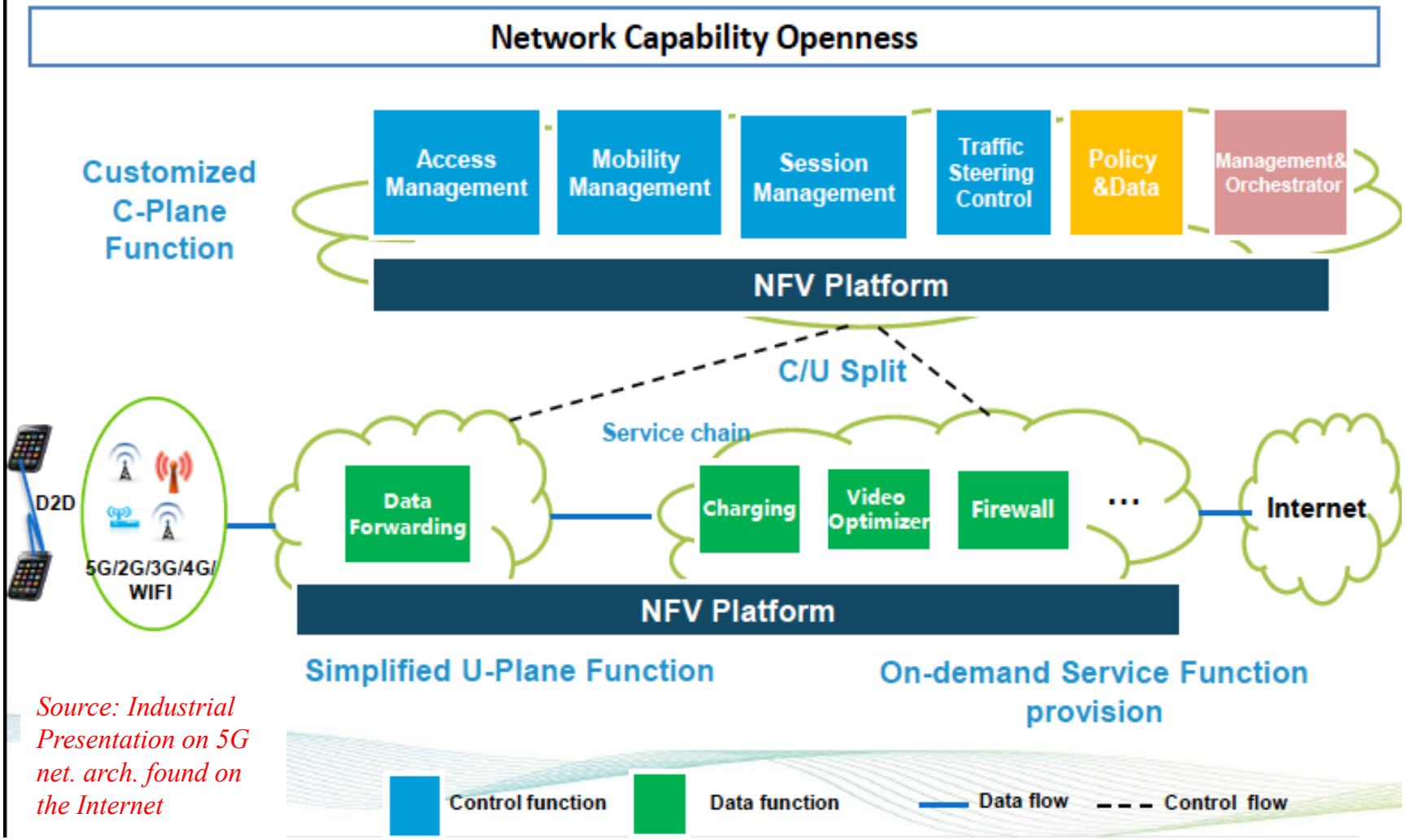
- We conducted a first measurement study of commercial 5G service deployment in US
  - Verizon, T-Mobile, Sprint in MPLS, CHI, ATL, …
- Main Take-Aways
  - Yes, 5G has the potential to support exciting new apps (up to 2 Gbps per mobile device, w/ 8 TCP connection)
  - Hugh implications on networking/edge computing: likely pushing bw bottlenecks inside the core; a lot of new challenges!



5G Is *Not* All About *New Radio*

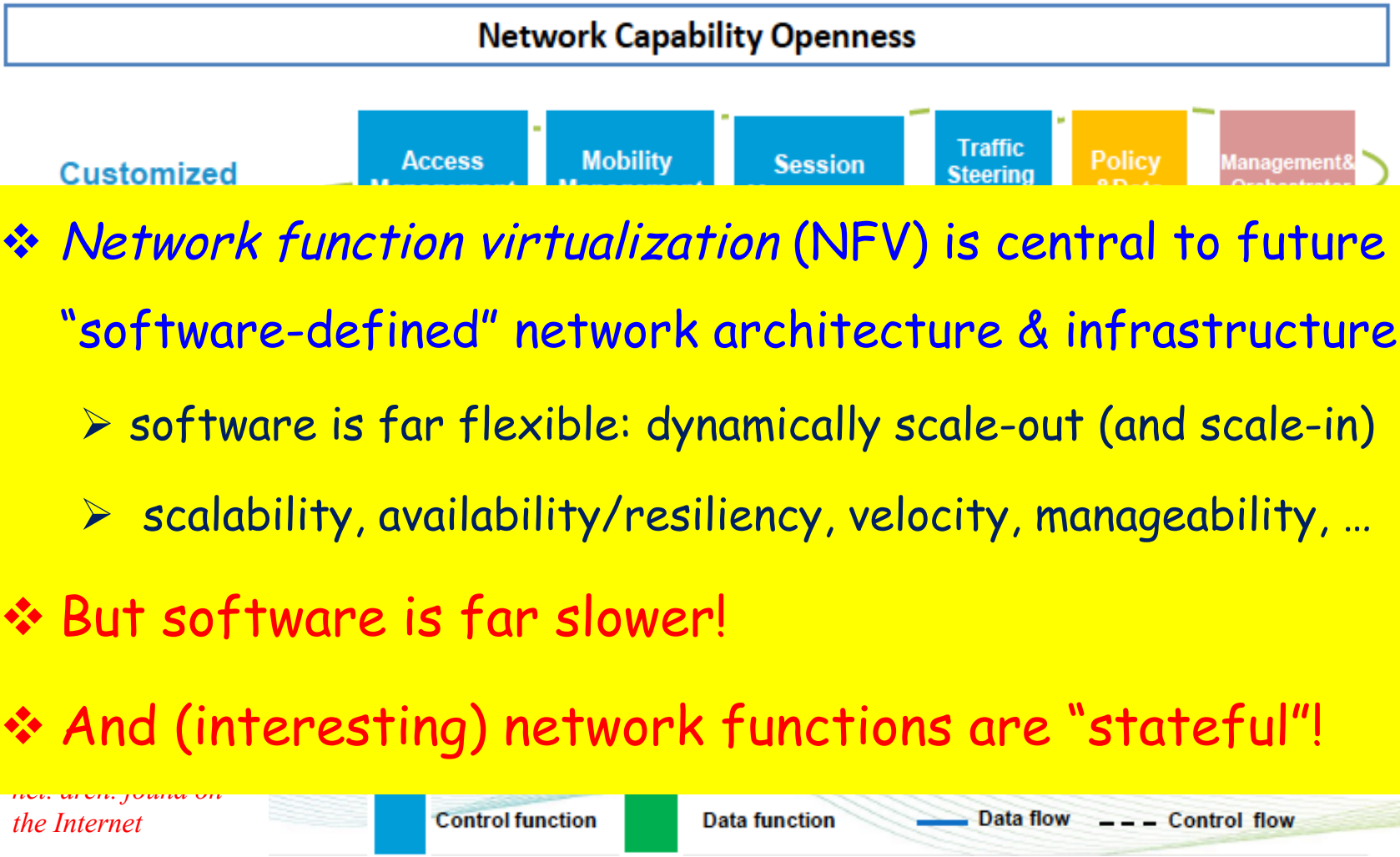
(Envisioned?) Commercial 5G Network Architecture



Source: Industrial Presentation on 5G net. arch. found on the Internet

5G Is *Not* All About *New Radio*

(Envisioned?) Commercial 5G Network Architecture

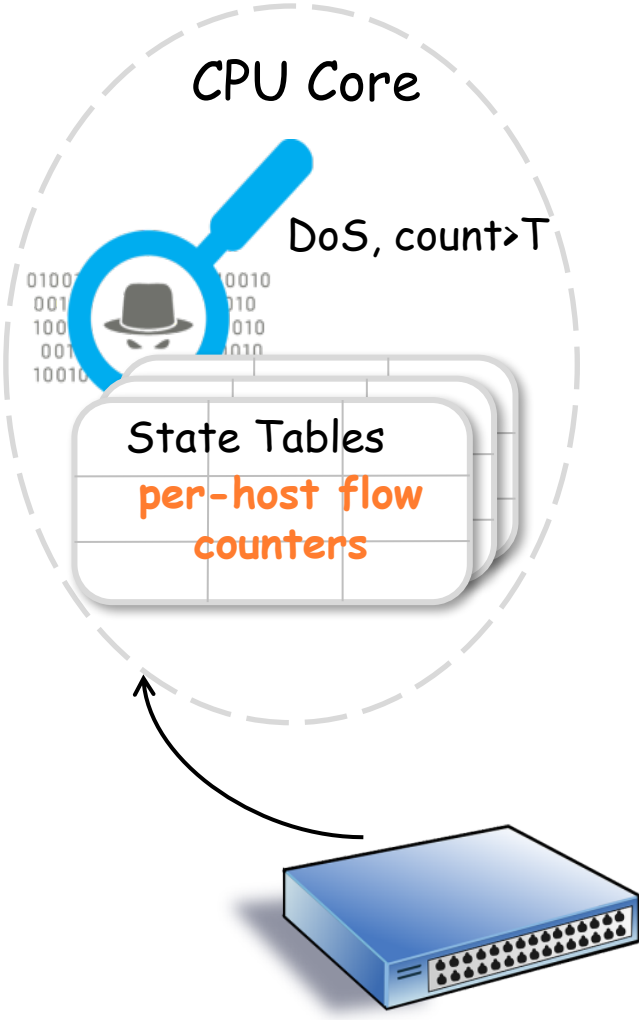


- *Network function virtualization* (NFV) is central to future "software-defined" network architecture & infrastructure
  - software is far flexible: dynamically scale-out (and scale-in)
  - scalability, availability/resiliency, velocity, manageability, …
- But software is far slower!
- And (interesting) network functions are "stateful"!

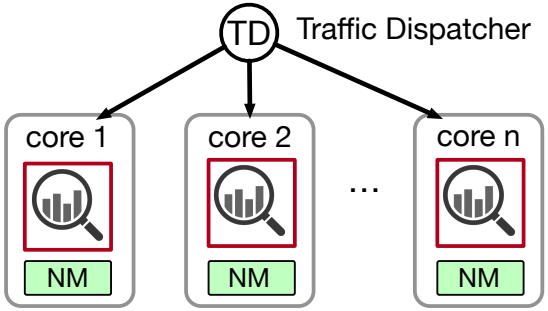
net. arch. found on the Internet

Scaling Out for Increased Throughput?

Network Monitoring (NM) NF:
count # of flows/packets sent by src IP



To meet increasing traffic demand,
Scaling out NM via *per-flow* traffic dispatching



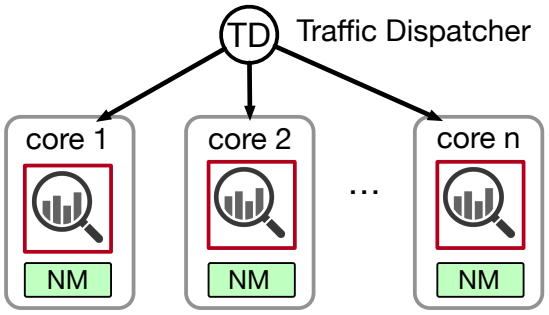
Scaling Out for Increased Throughput?

- Allocating more cores by running more NF instances when NF instances *share "state"* may not always help!

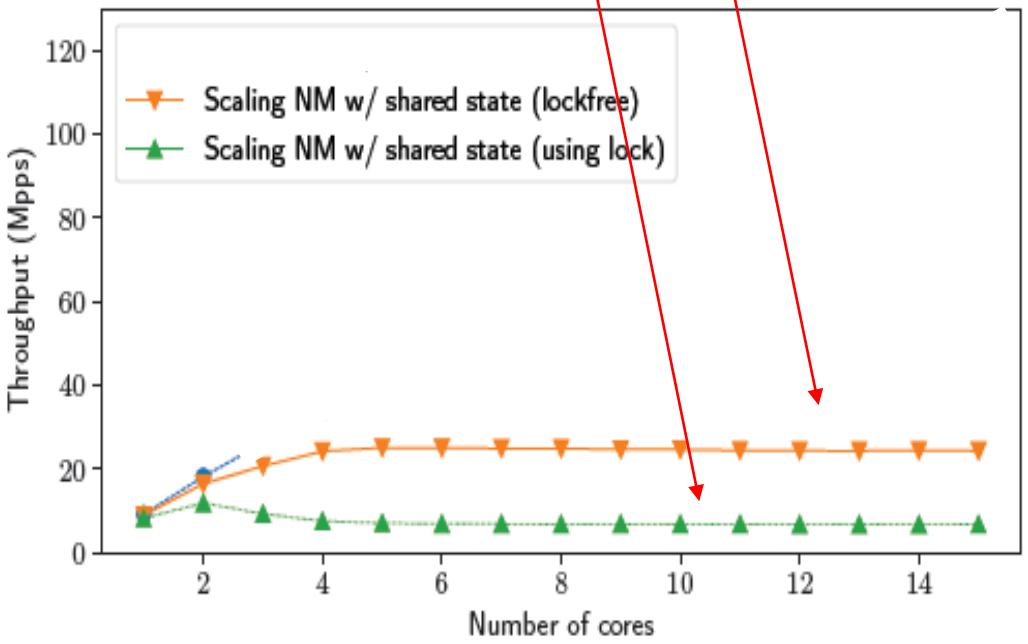
Scaling out NFV performance via multiple cores no longer linear!

In some worst cases, more cores can even hurt performance

Scaling out NM via *per-flow* traffic dispatching



Shared L3/DRAM



Scaling Out for Increased Throughput?

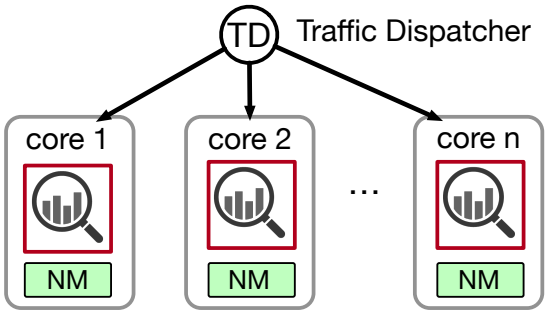
Can't always add more servers to meet throughput demands

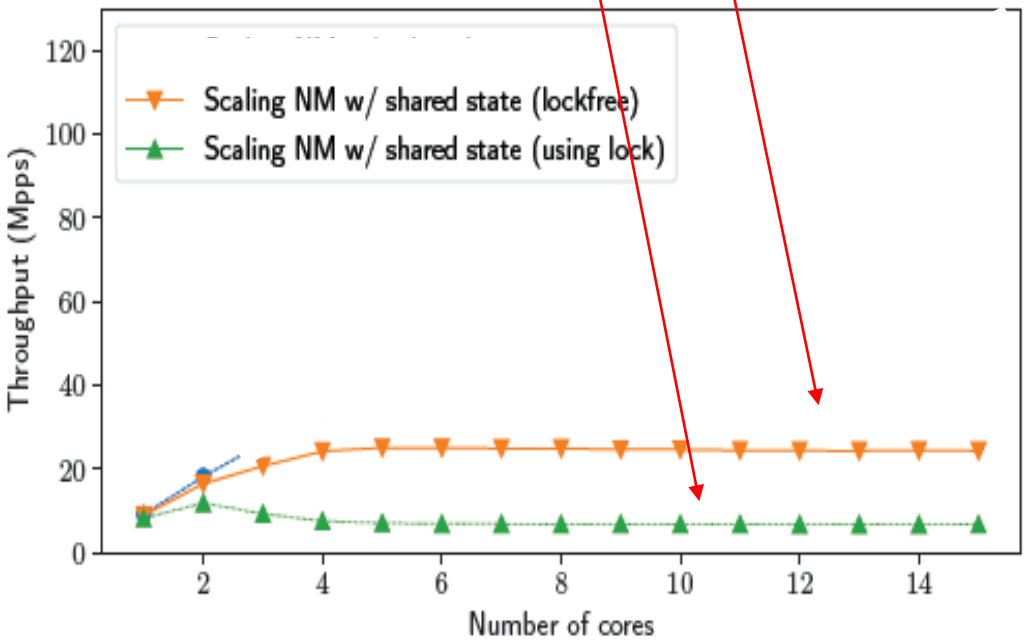
- **edge cloud facility has limited real estate, power & other constraints**
- **need to increase per-core/server software pkt processing capability**

Scaling out NFV performance via multiple cores no longer linear!

In some worst cases, more cores can even hurt performance

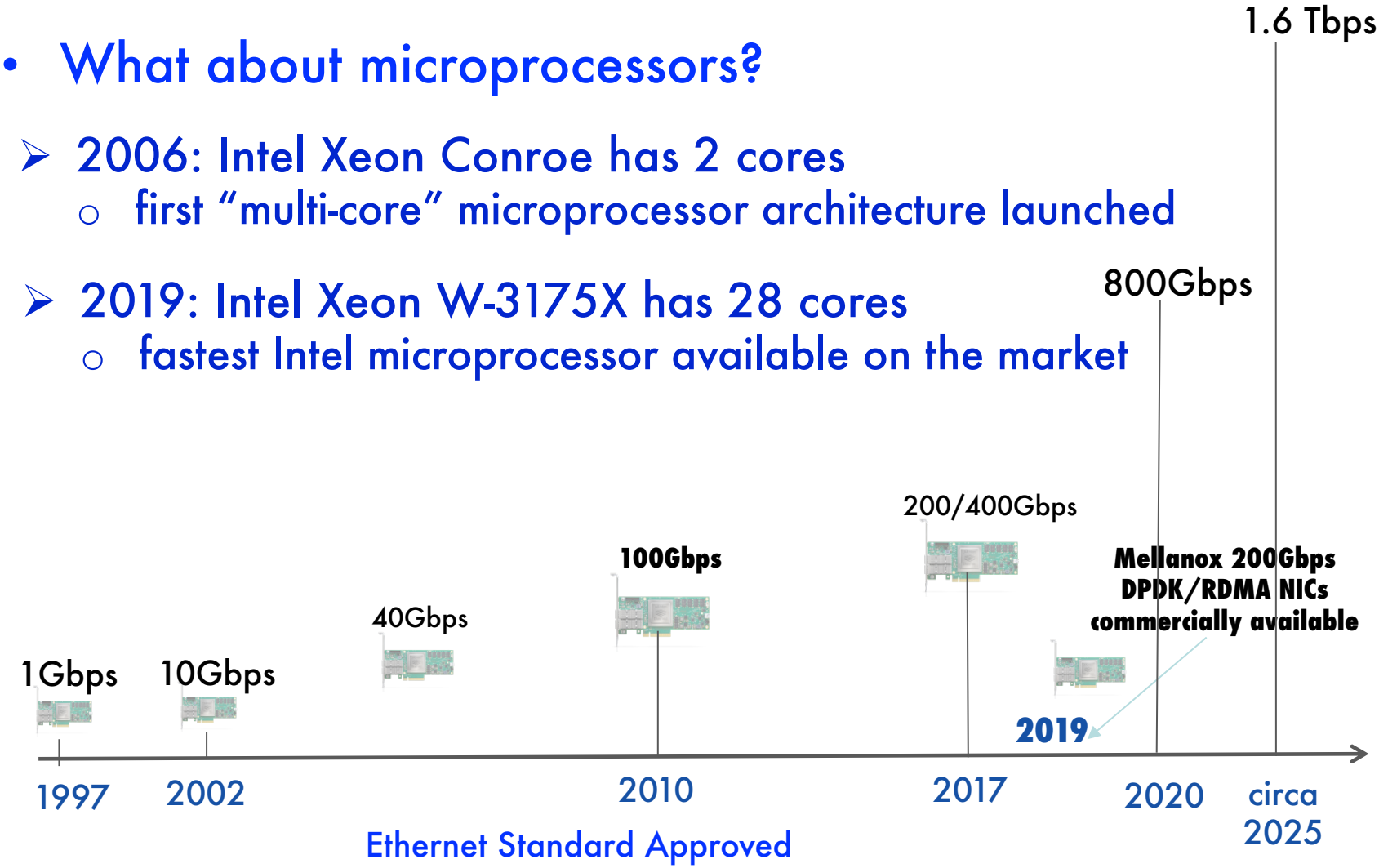
Scaling out NM via *per-flow* traffic dispatching





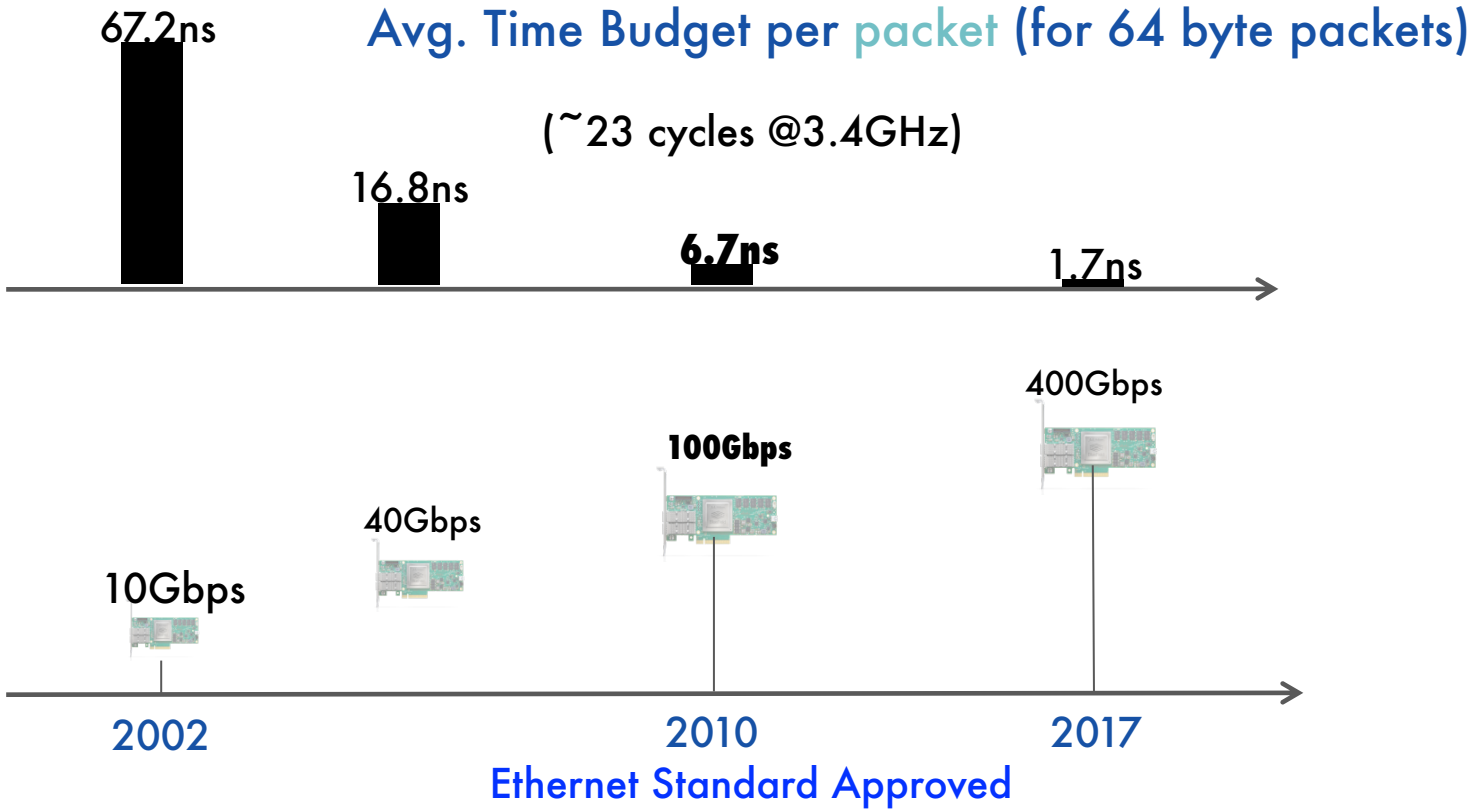
Widening Gap between Net BW & CPU

- What about microprocessors?
  - 2006: Intel Xeon Conroe has 2 cores
    - first "multi-core" microprocessor architecture launched
  - 2019: Intel Xeon W-3175X has 28 cores
    - fastest Intel microprocessor available on the market

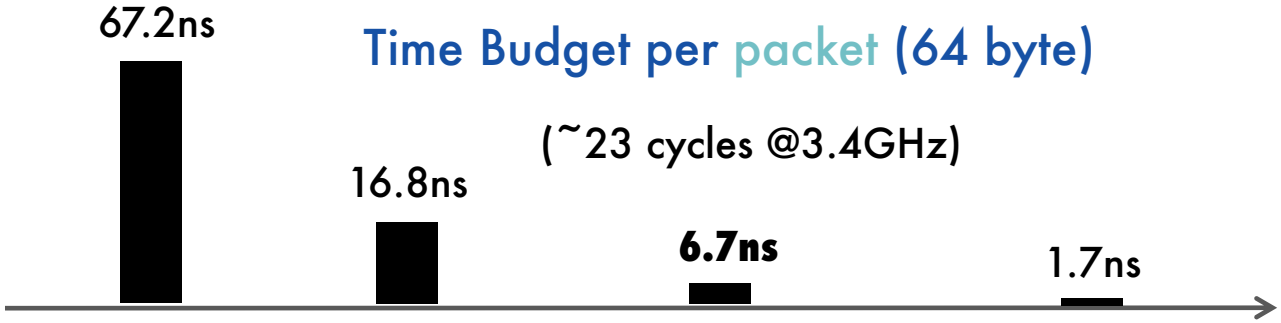


Widening Gap between Net BW & CPU

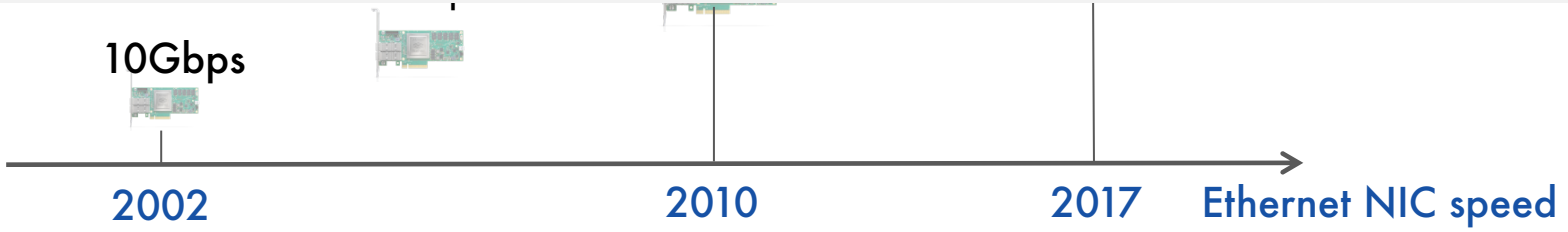
- Why attaining line speed for NFV is challenging



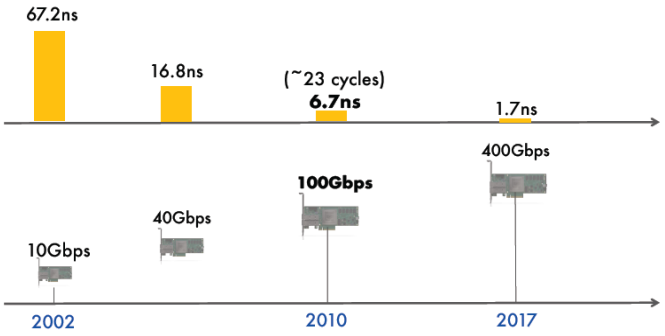
Challenges in Keeping up with Growing Line speed



How could an NFV execution target (multi-core server) support such *tight* time budget for per-pkt processing?

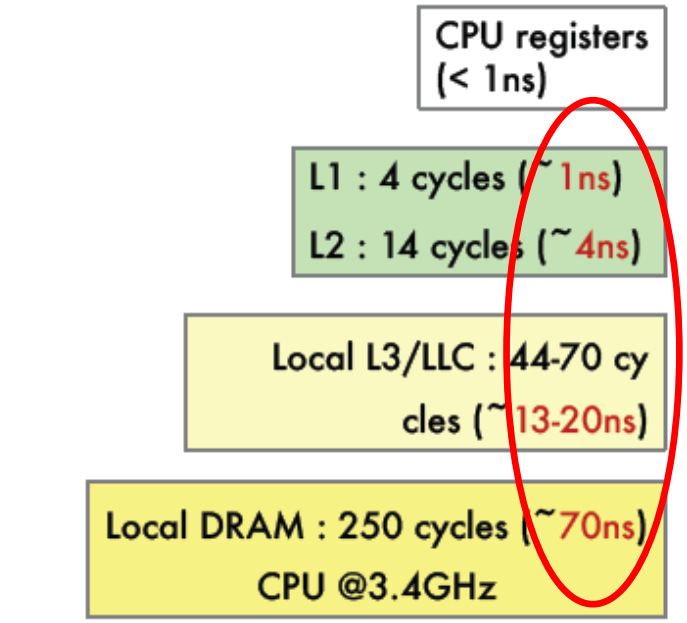


Sw Packet Processing & Multi-Core Servers

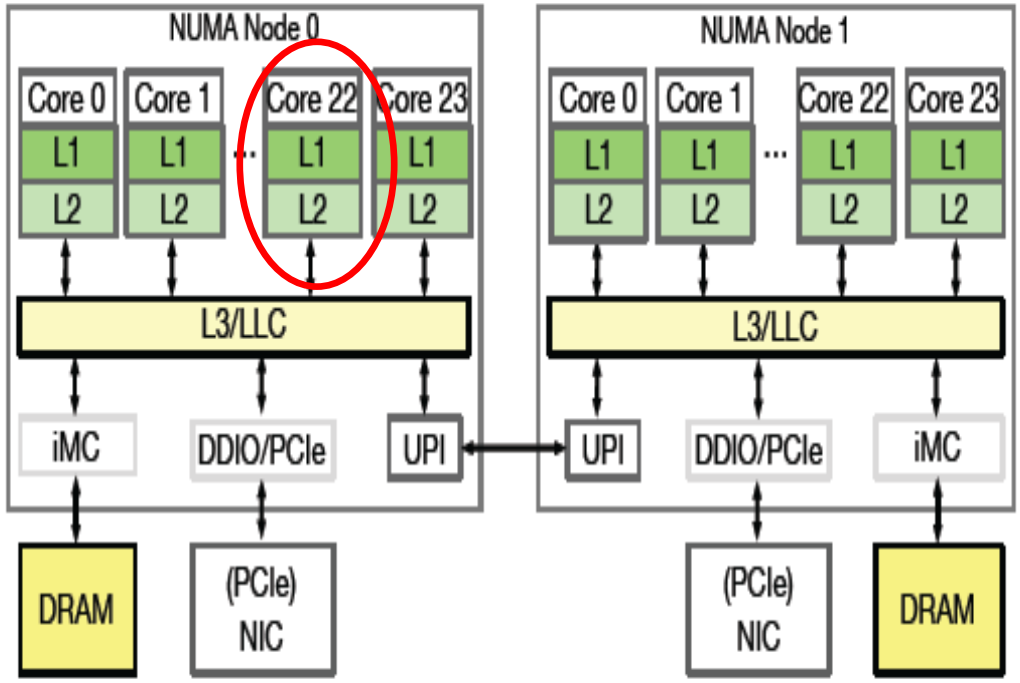


Ensuring most NF operations are L1/L2 bound

is important for 100Gbps line speed



Memory hierarchy of NUMA servers



Intel(R) Xeon(R) Platinum 8168, dual CPU sockets w/ 24 cores each, CPU @2.7GHz clocked at 3.4GHz,

NFV Research & Experimental Infra.

- Network function virtualization (NFV) is central to future "software-defined" network architecture & infrastructure
  - software far flexibility, dynamically scale-out (and scale-in)
  - scalability, availability/resiliency, velocity, manageability, …
- But software is far slower!
- And (interesting) network functions are "stateful"!
- **A lot of challenging research issues!**
- **Benchmarking & experimentations (instead of hype)**
- **Collaborative NFV research infrastructure critical!**

Thank You!