UNICAMP

Institute of Computing

Computer Networks Laboratory

INCT InterSCity

Multi-Tier Edge-to-Cloud Architecture for Adaptive Video Delivery

*Roger Immich, Leandro Villas, **Luiz Bittencourt**, Edmundo Madeira*

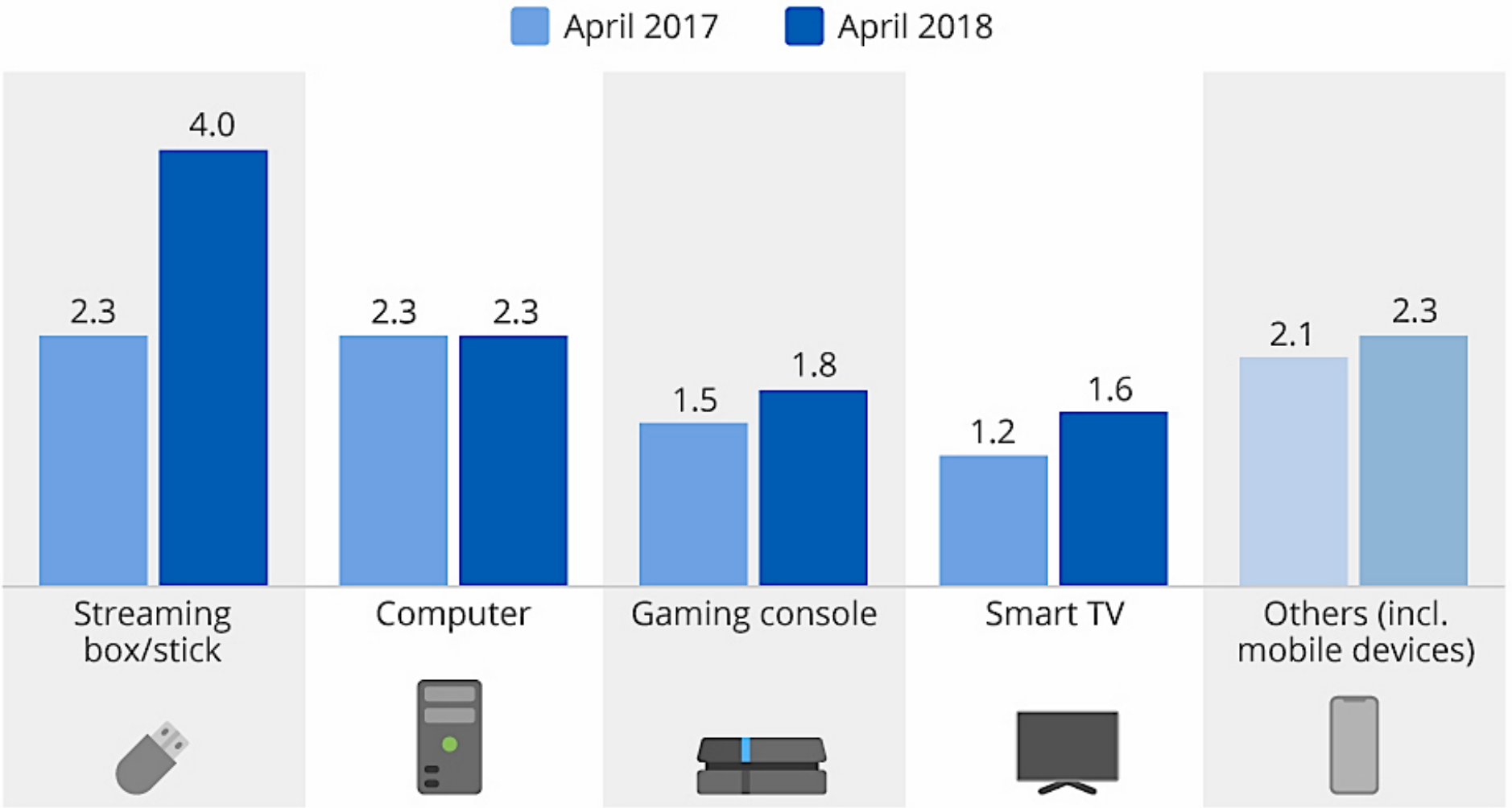
FiCloud - August 2019, Istanbul, Turkey

Outline

- Introduction / Motivation
- Edge computing
- Microservices
- Multi-tier video delivery architecture
- Experiment setup and assessment
- Summary

Streaming Devices Are America's Biggest Traffic Hogs

Aggregate amount of data received by all wi-fi households in the U.S. (in billion gigabytes)

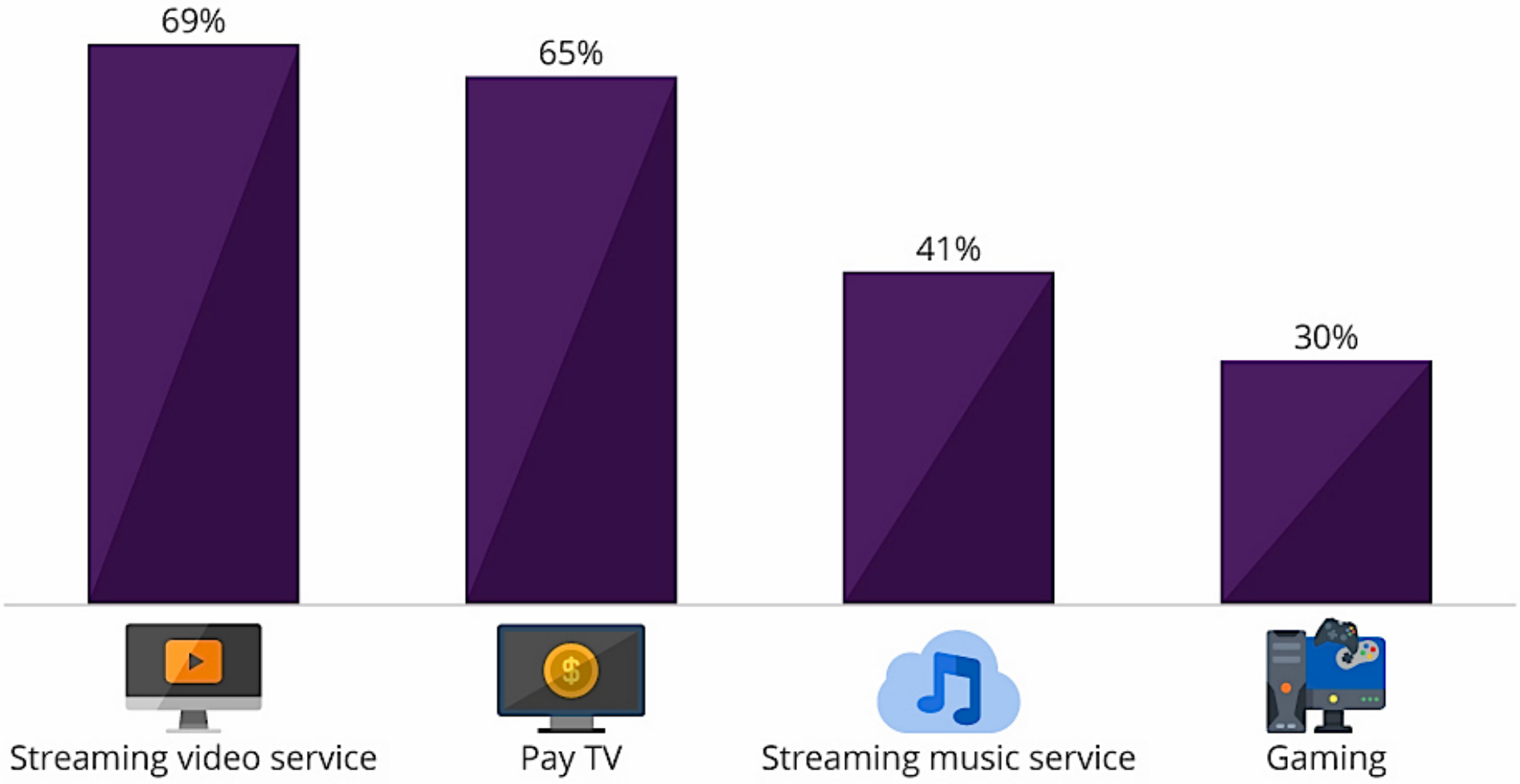


@StatistaCharts Source: comScore

statista

Streaming Services Overtake Pay TV In The U.S.

Share of U.S. households with the following subscriptions in 2018



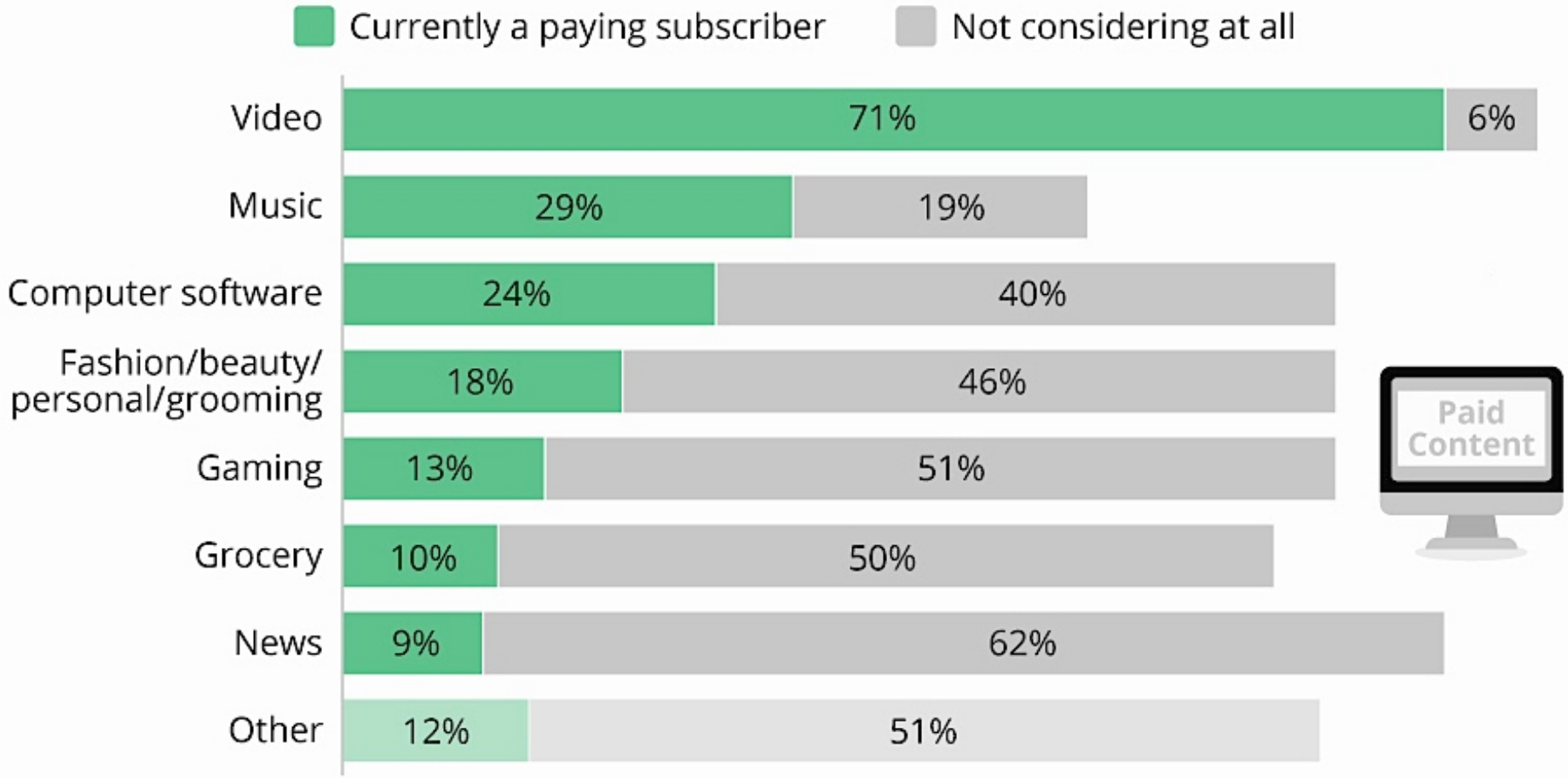
n=2,003 U.S. consumers (Dec 2018 - Feb 2019)
Source: Deloitte

@StatistaCharts

statista

Video Wins Subscribers.

Most popular paid subscription service categories in the U.S. 2018*



* 1,500 respondents surveyed in the United States in June 2018. Subscription services defined as previously pay-per product/service companies that now charge customers a subscription price to have recurring access to their products or services.

@StatistaCharts

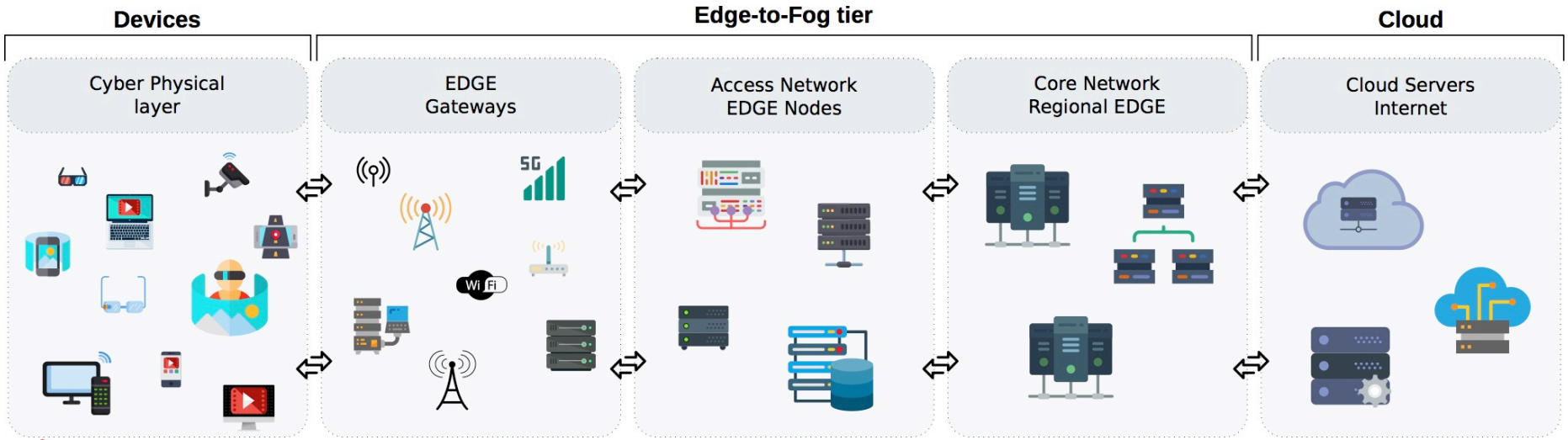
Sources: eMarketer, Pollfish, Fetch

statista

What is edge computing?

Cloud principles applied at the network edge

- Virtualization (CPU, Storage, Networking)
- On Demand
- API driven
- Automated LCM
- Commodity hardware



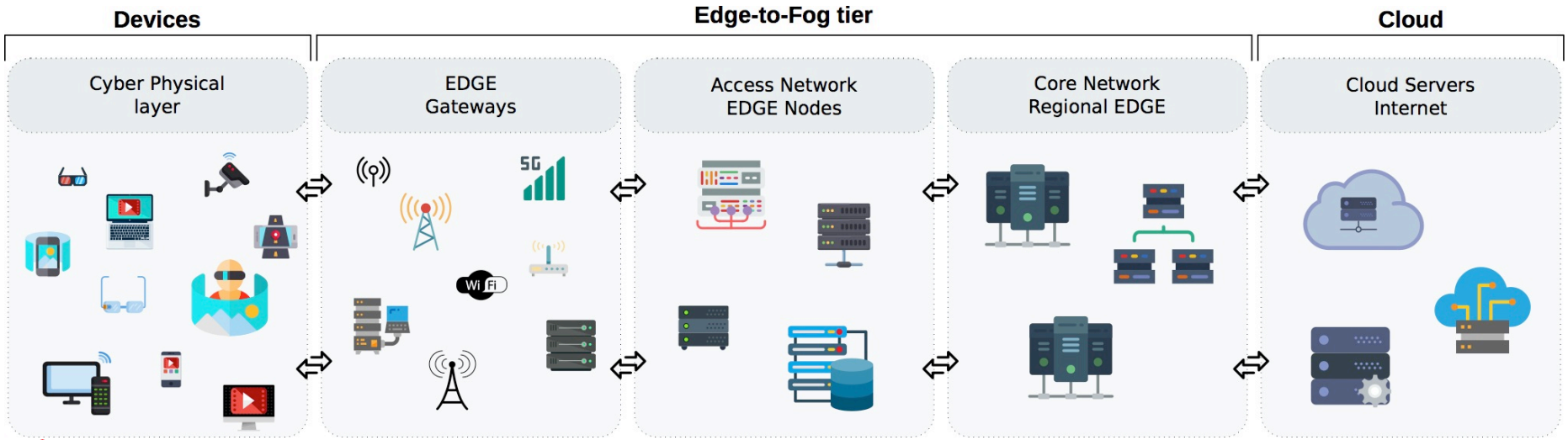
What is edge computing?

Cloud principles applied at the network edge

- Virtualization (CPU, Storage, Networking)
- On Demand
- API driven
- Automated LCM
- Commodity hardware

Convergence of IT and telecom networking

- Allows network operators to open up their networks to new opportunities and value chains



What is edge computing?

Cloud principles applied at the network edge

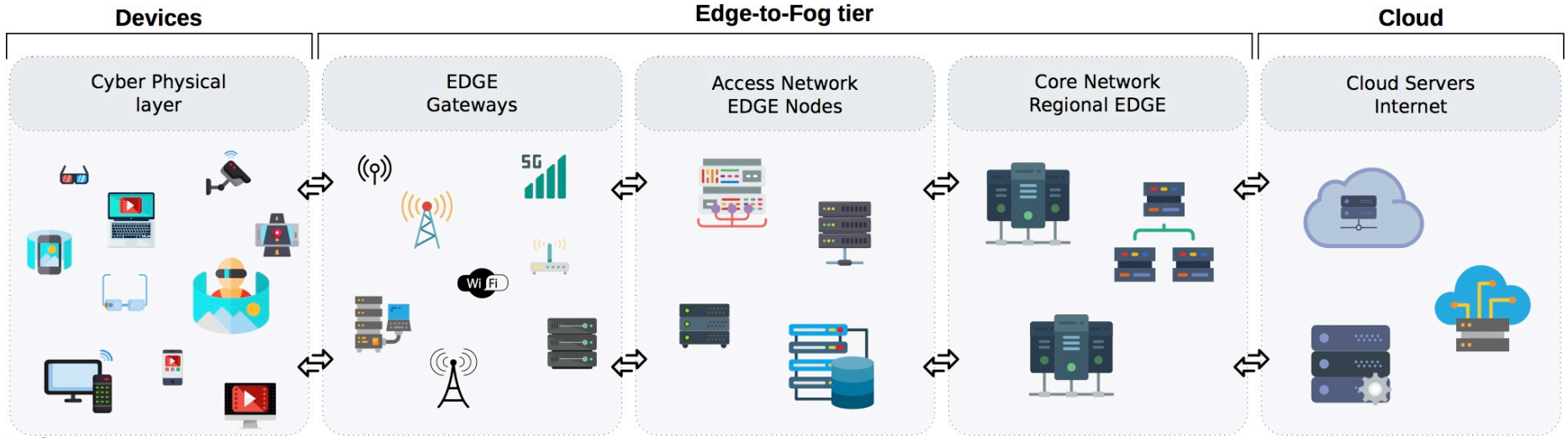
- Virtualization (CPU, Storage, Networking)
- On Demand
- API driven
- Automated LCM
- Commodity hardware

Convergence of IT and telecom networking

- Allows network operators to open up their networks to new opportunities and value chains

New, Innovate, Immersive applications

- Opportunity for:
- Tailored apps to local conditions
- Provide contextualized services
- Low latency, high bandwidth guarantees



Cloud and Edge Computing

- **Cloud capabilities at the edge of the network**
- Takes advantage of a **shared data center** infrastructure and the **economy of scale** to reduce costs
- Emerging application requirements: **low latency**, **high bandwidth**
  - Smart cities, Video Services, IoT, Tactile Internet, Augmented Reality, etc...

Multi-access edge computing (MEC)

- **Cloud-computing** capabilities within the **RAN** in **close proximity** to mobile subscribers
- Accelerates content, services, and applications so increasing **responsiveness**
- RAN edge offers:
  - **Ultra-low latency** and **high-bandwidth**
  - Direct access to **real-time radio network information** (subscriber location, cell load, link quality, etc.)
- Essential to offer **context-related services**

Microservices

“**service-oriented architecture** *composed of* **loosely coupled elements** *that have* **bounded contexts**”

source: Adrian Cockcroft (Cloud Architect at Netflix)

Microservices

"**service-oriented architecture** *composed of* **loosely coupled elements** *that have* **bounded contexts**"

Services communicate with each other over the network

source: Adrian Cockcroft (Cloud Architect at Netflix)

Microservices

"**service-oriented architecture** *composed of* **loosely coupled elements** *that have* **bounded contexts**"

Services communicate with each other over the network

You can update the services independently; updating one service doesn't require changing any other services

source: Adrian Cockcroft (Cloud Architect at Netflix)

Microservices

"**service-oriented architecture** *composed of* **loosely coupled elements** *that have* **bounded contexts**"

Services communicate with each other over the network

You can update the services independently; updating one service doesn't require changing any other services

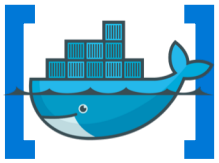
Self-contained; update the code without knowing anything about the internals of other microservices

source: Adrian Cockcroft (Cloud Architect at Netflix)

Multi-tier video delivery architecture

- Aims to advance the idea of multi-tier video delivery using off-the-shelf open-source tools
- Relying upon technologies such as network slices and microservice placement
- Proof-of-concept testbed and real video sequences
- The main goal is to prove that it is possible to build a real multi-tier environment to improve video delivery quality

Experiment setup



Docker

Resource management (container)

Kubernetes

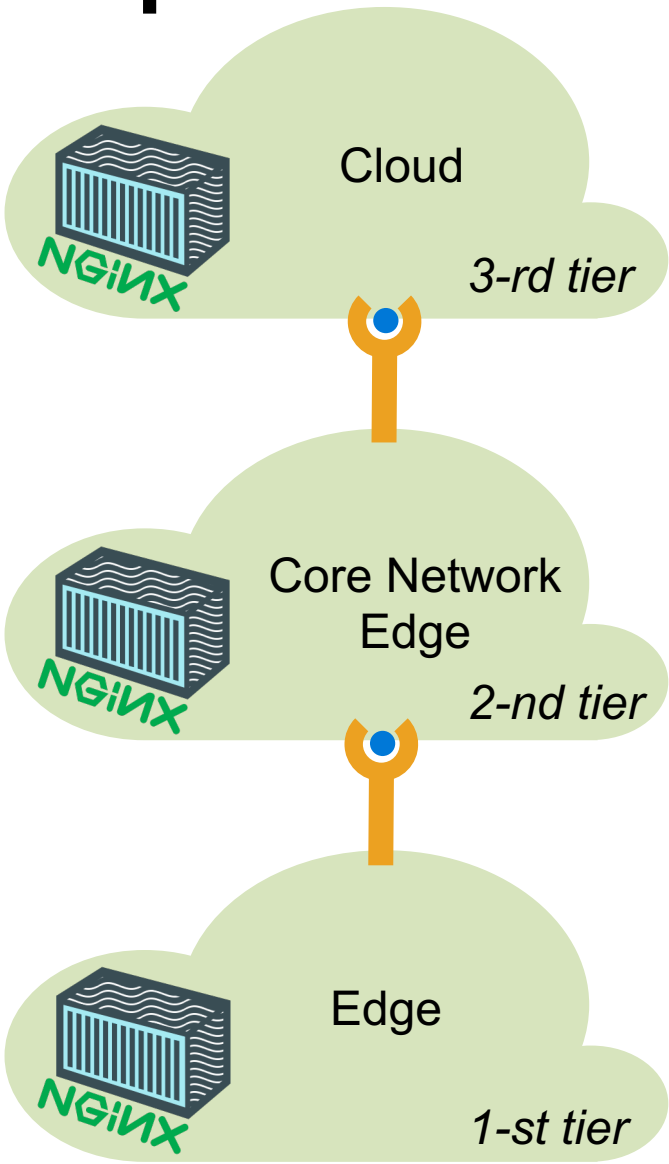
Container orchestration

Juju

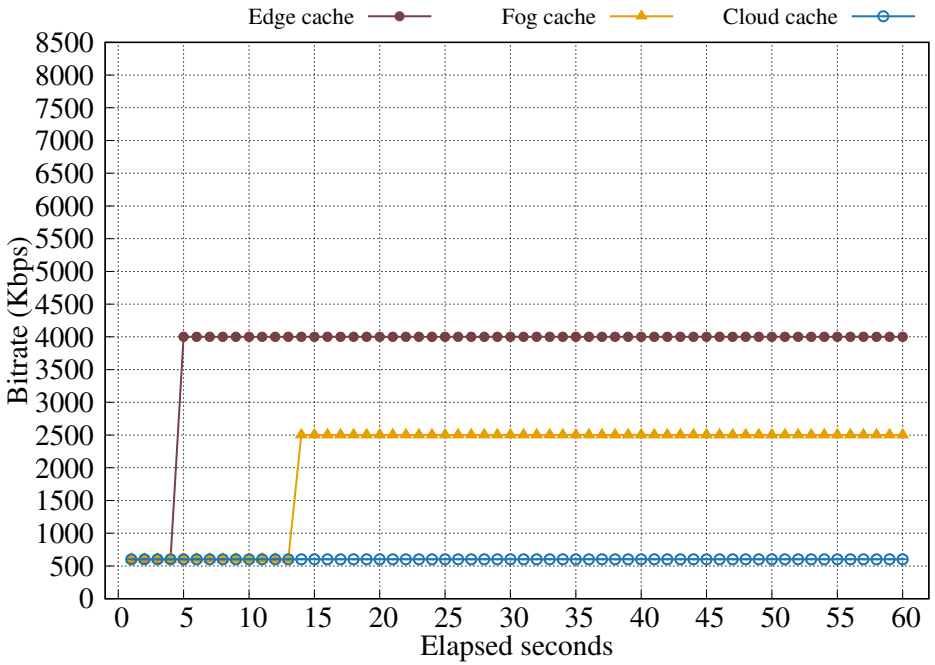
Application management

Reference Client 2.9.0

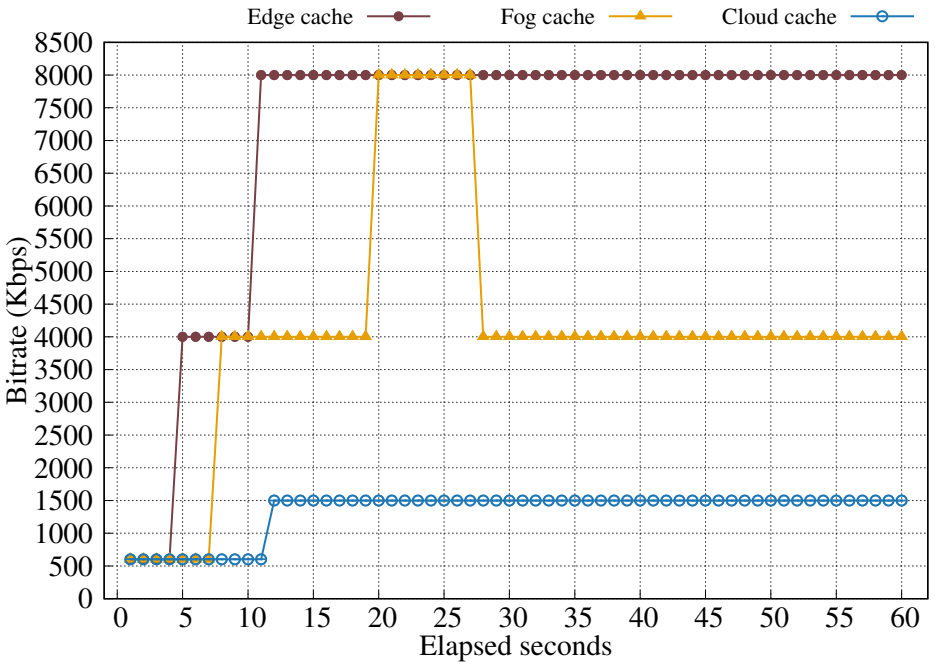
mpeg-DASH



Bitrate assessment

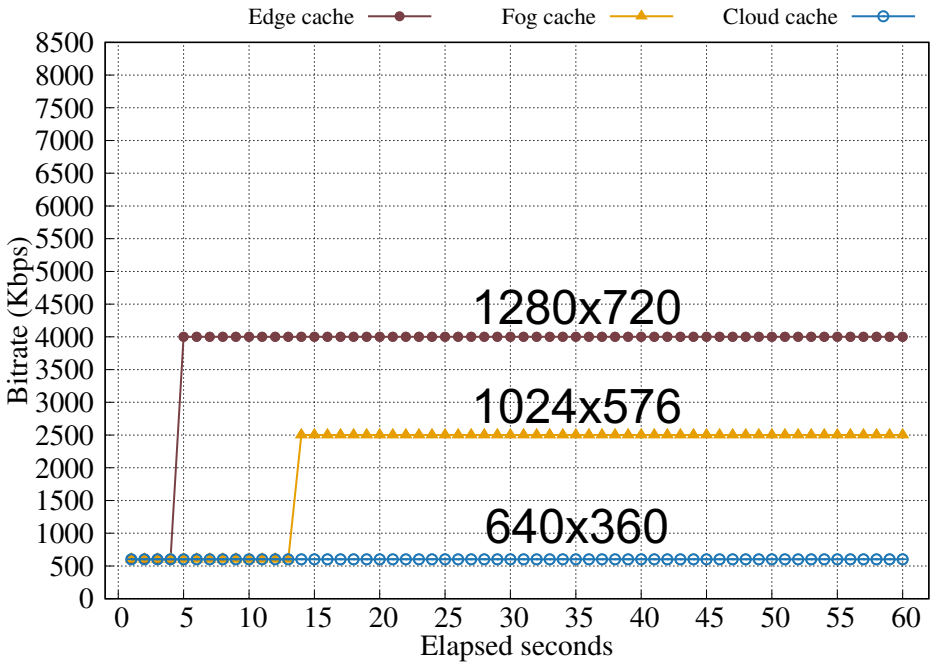


Delivered bitrate **without** network slices

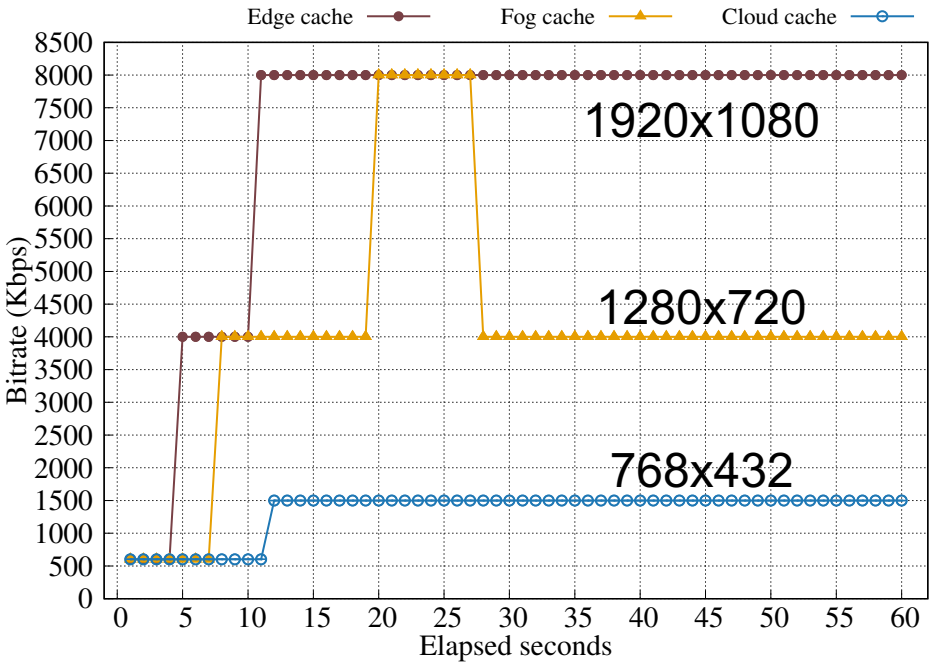


Delivered bitrate **with** network slices

Bitrate assessment

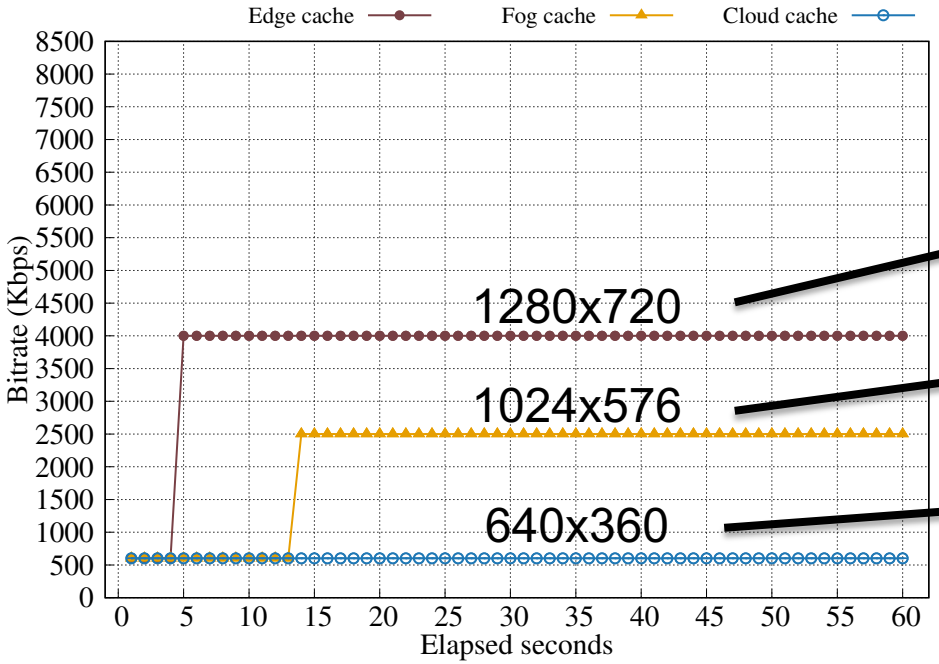


Delivered bitrate **without** network slices

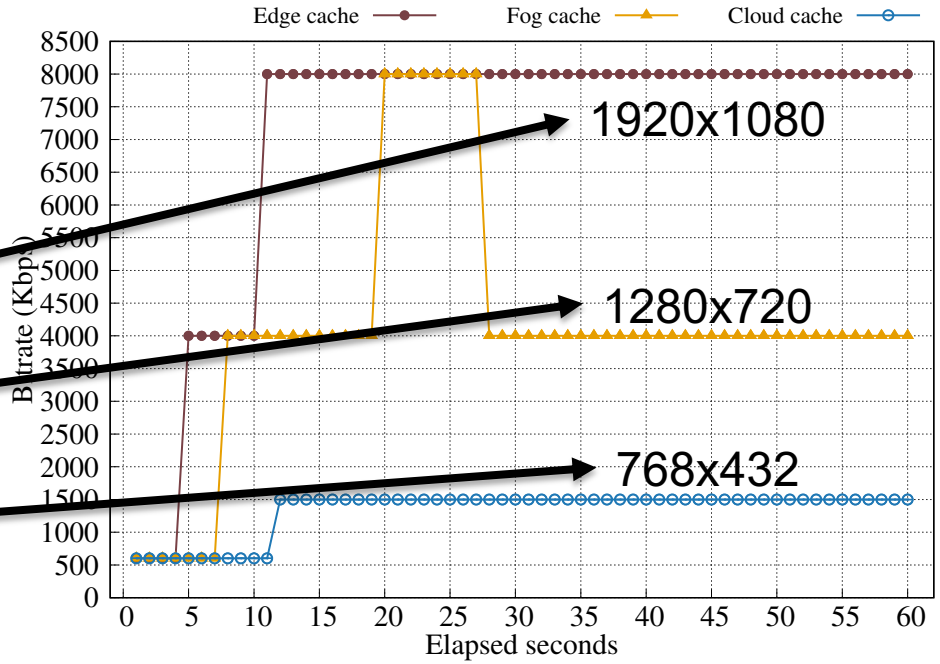


Delivered bitrate **with** network slices

Bitrate assessment

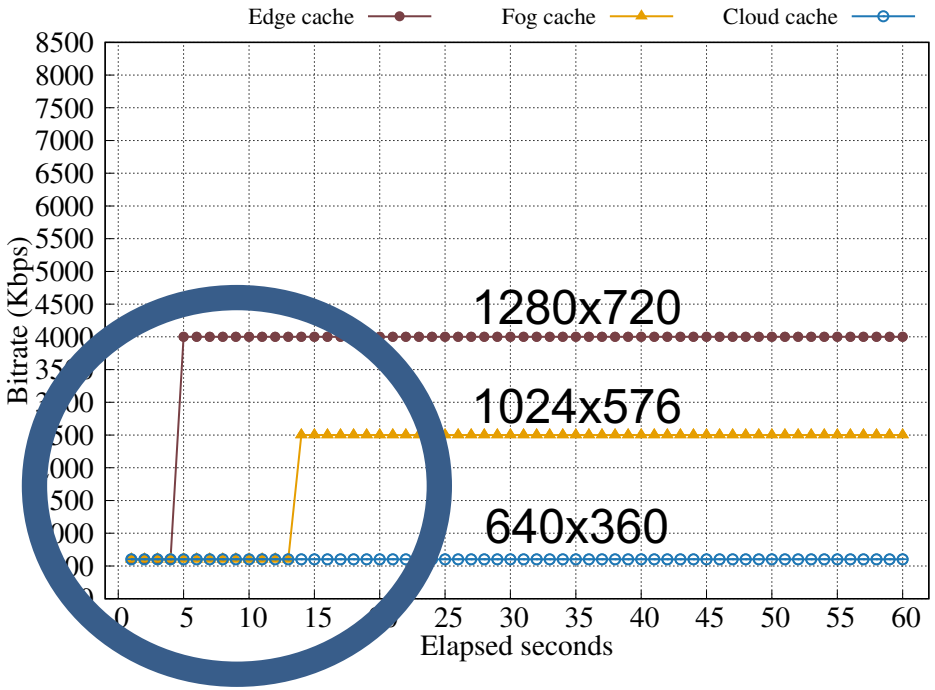


Delivered bitrate **without** network slices

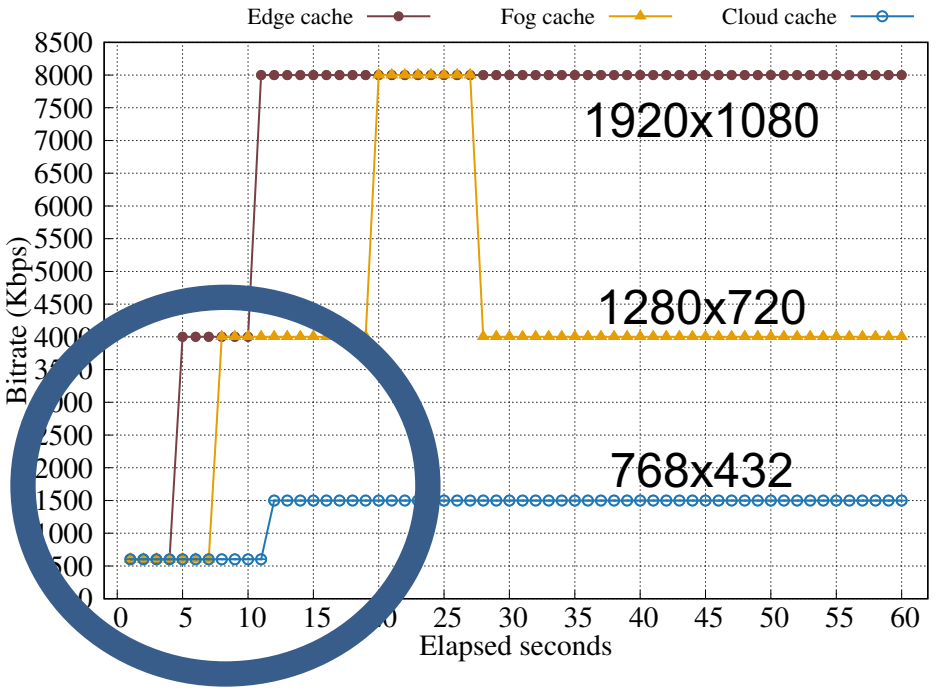


Delivered bitrate **with** network slices

Bitrate assessment



Delivered bitrate **without** network slices

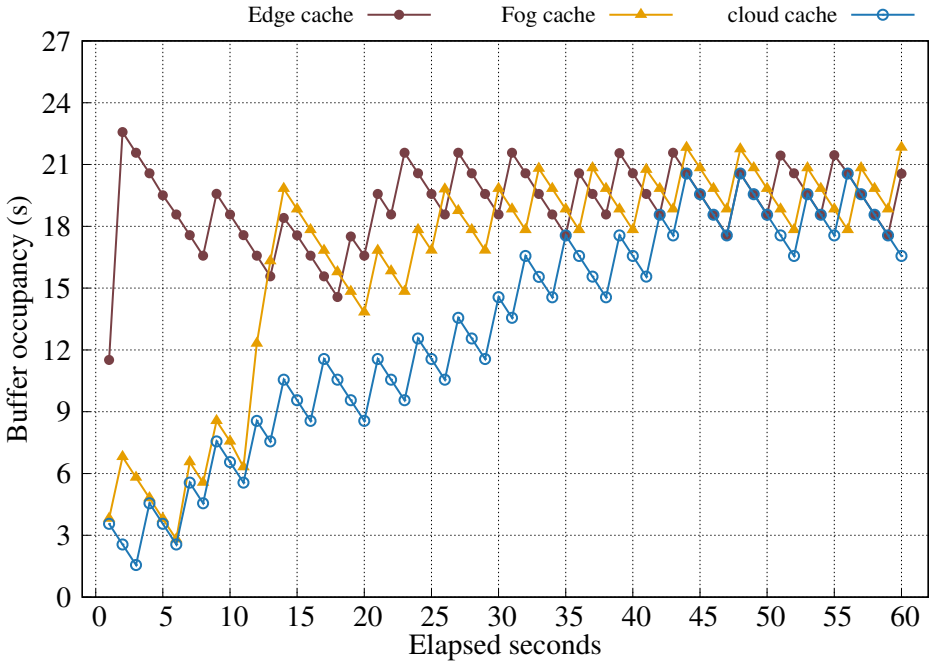


Delivered bitrate **with** network slices

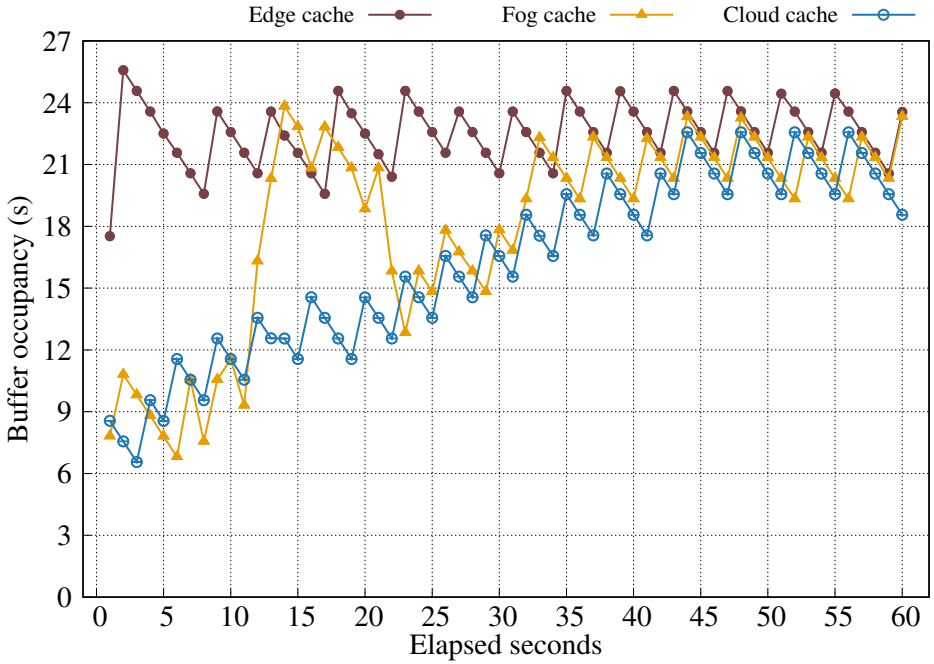
Slice: lower time taken to increase resolution

Buffer occupancy assessment

- Healthy buffer hides delay/latency/jitter

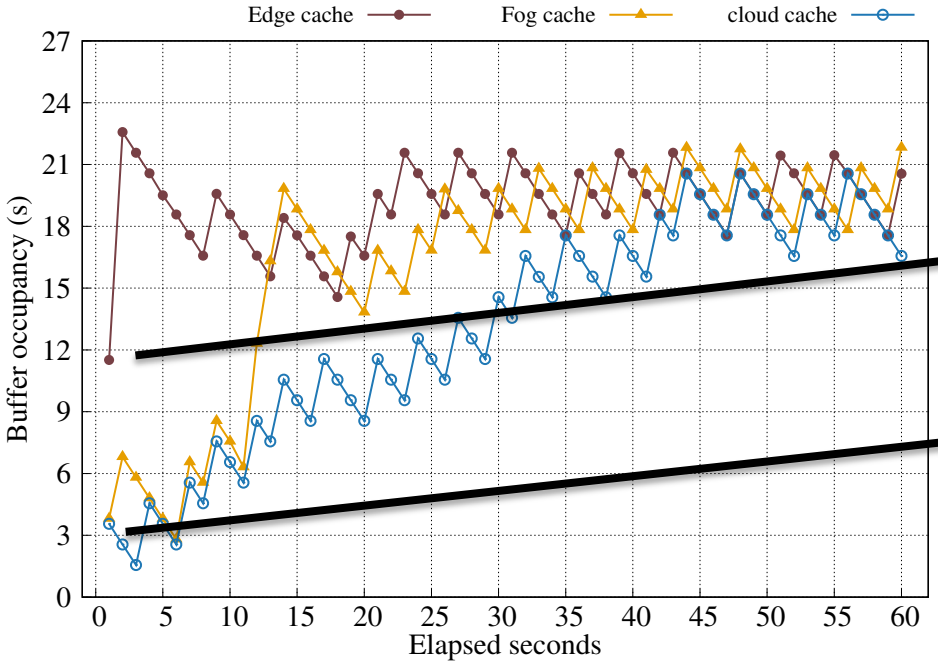


Buffer occupancy **without** network slices

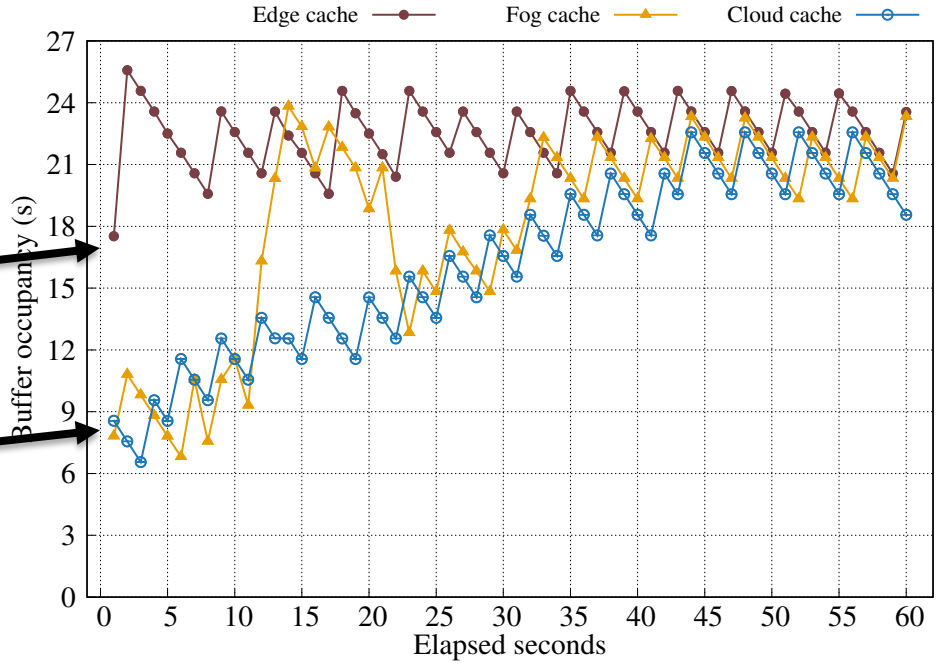


Buffer occupancy **with** network slices

Buffer occupancy assessment



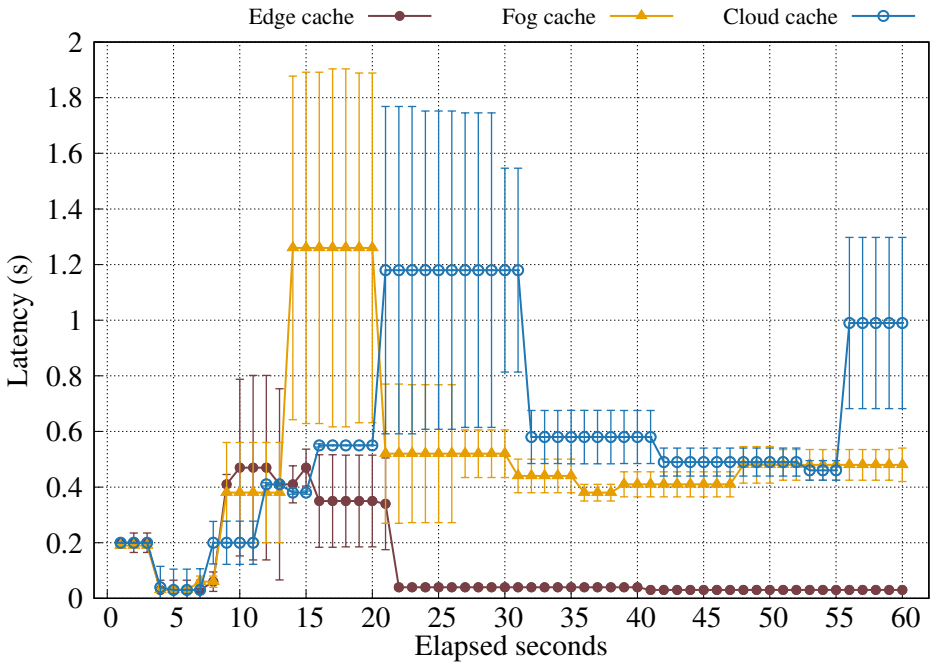
Buffer occupancy **without** network slices



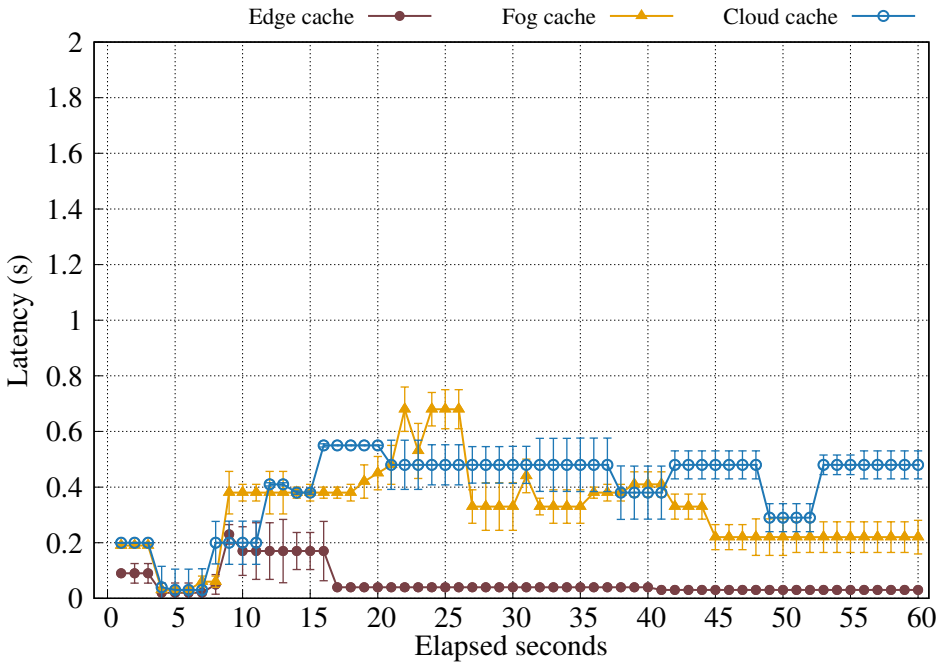
Buffer occupancy **with** network slices

Slices: Optimal buffer level achieved earlier

Latency assessment

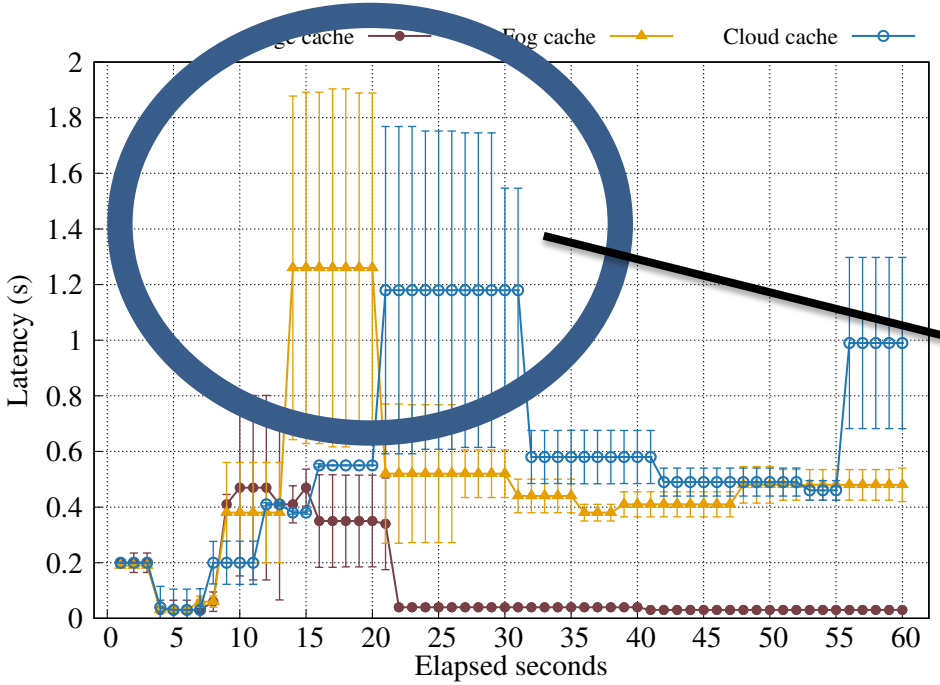


Latency **without** network slices

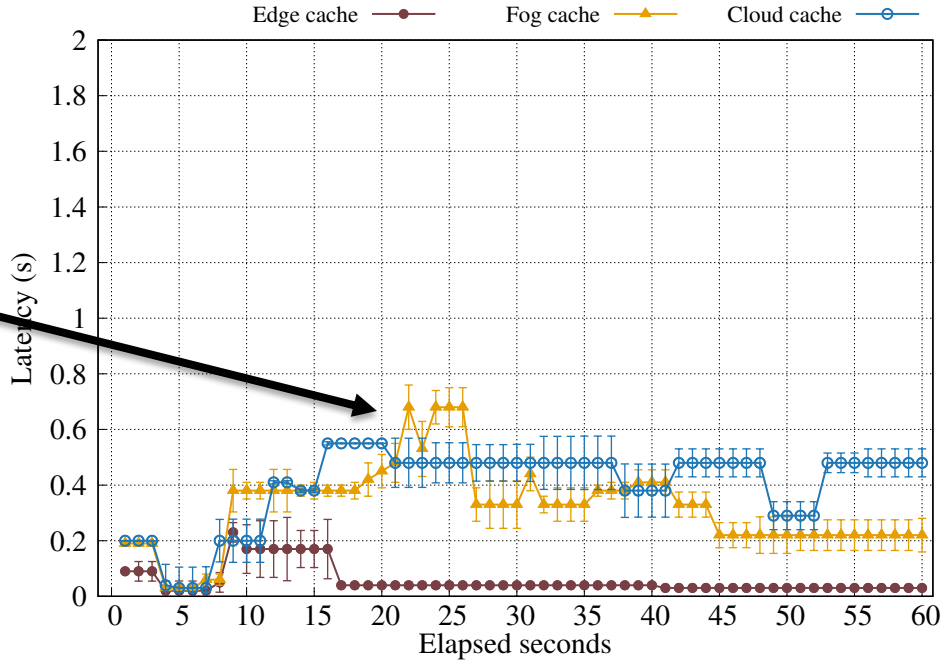


Latency **with** network slices

Latency assessment



Latency **without** network slices



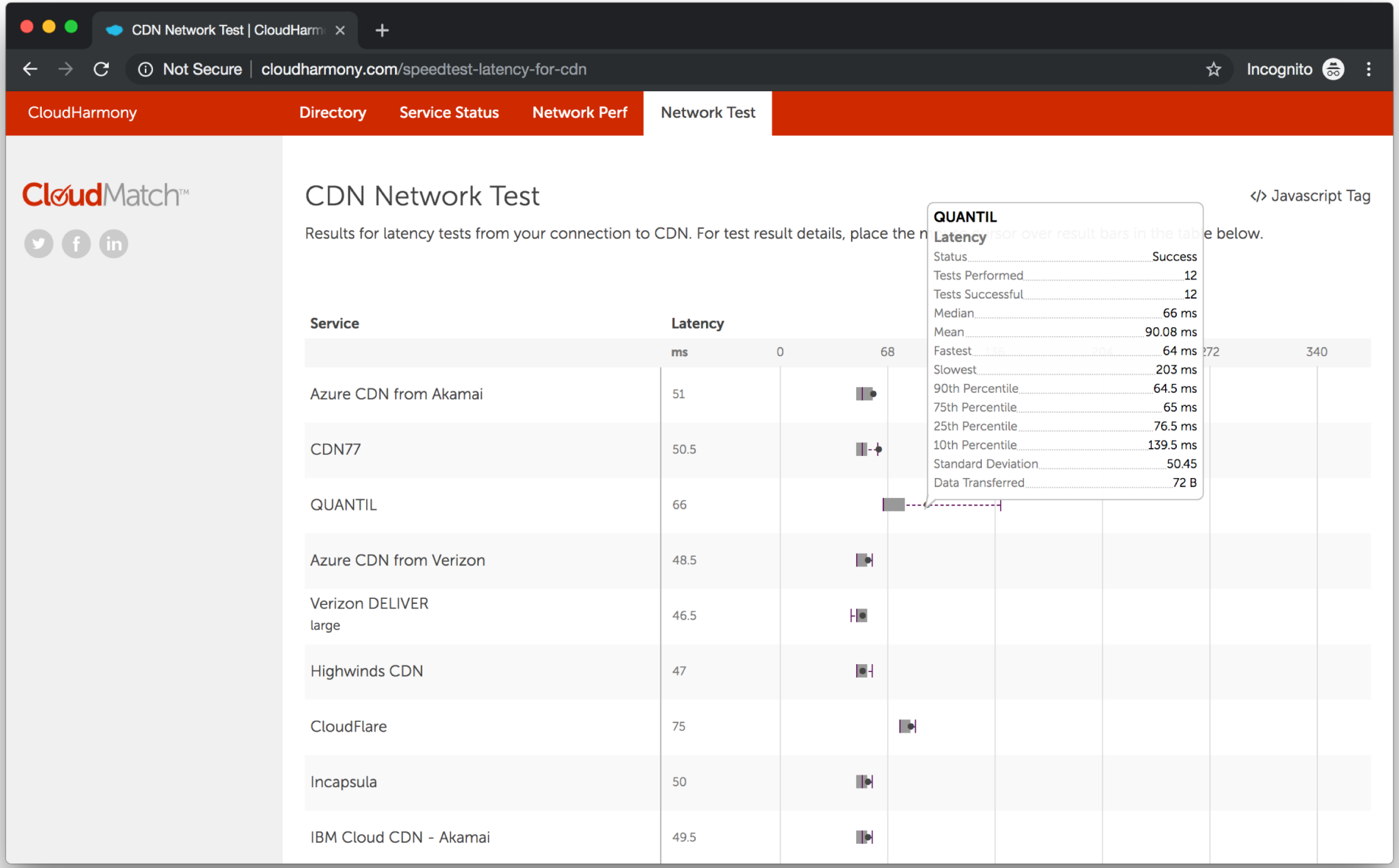
Latency **with** network slices

Slices: latency and standard deviation are considerably smaller

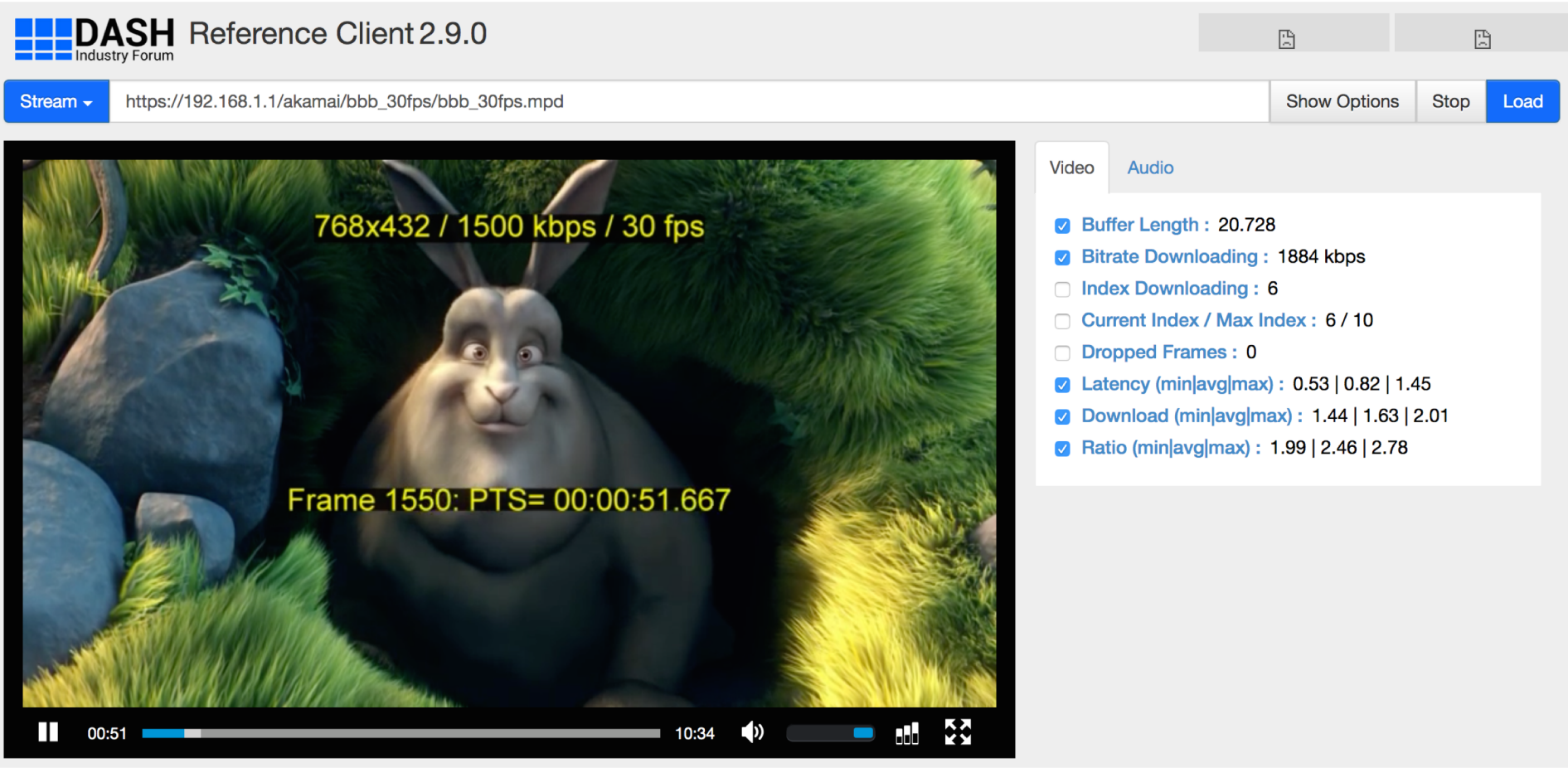
Summary

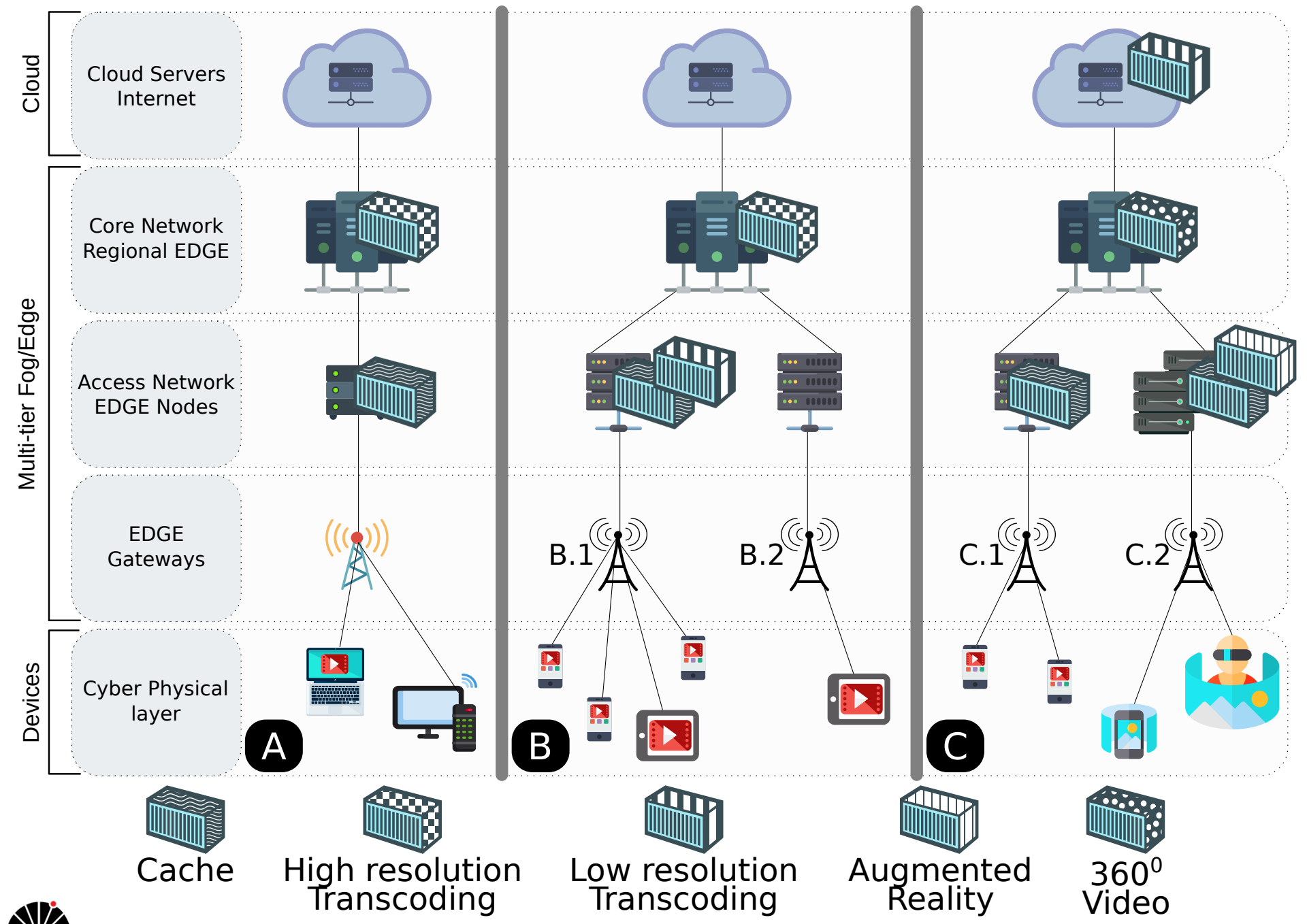
- The combination of 5G, slices, MEC, containers, and microservices provides a **highly distributed computing environment**
- Applications, services, store and process content in **close proximity** to the **users**
- As future work, applications will benefit from real-time **radio and network** information
- Provide **Personalized** and **contextualized** experience to the subscribers

Thank You !


CDN Network Test | CloudHarm
Not Secure | cloudharmony.com/speedtest-latency-for-cdn
Incognito
CloudHarmony
Directory
Service Status
Network Perf
Network Test
CloudMatch™
CDN Network Test
Javascript Tag
Results for latency tests from your connection to CDN. For test result details, place the n
e below.
QUANTIL
Latency
Status Success
Tests Performed 12
Tests Successful 12
Median 66 ms
Mean 90.08 ms
Fastest 64 ms
Slowest 203 ms
90th Percentile 64.5 ms
75th Percentile 65 ms
25th Percentile 76.5 ms
10th Percentile 139.5 ms
Standard Deviation 50.45
Data Transferred 72 B
Service
Latency
ms
0
68
272
340
Azure CDN from Akamai 51
CDN77 50.5
QUANTIL 66
Azure CDN from Verizon 48.5
Verizon DELIVER large 46.5
Highwinds CDN 47
CloudFlare 75
Incapsula 50
IBM Cloud CDN - Akamai 49.5

PARAMETERS	VALUE
Display sizes	320x180 up to 3840x2160
Frame rate	30 fps
Aspect ratio	16:9
Video mimeType	video/mp4
MPEG4 video file	m4v
Audio mimeType	audio/mp4
audioSamplingRate	48 kHz
MPEG4 audio file	m4a
Dash Player	Reference Client 2.9.0
Dash Schedule While Paused	Not selected
Dash Allow Local Storage	Not selected
Dash Low Latency Mode	Not selected
Dash Jump Small Gap	Selected
Dash Fast Switching ABR	Selected
Dash Fast Switching Strategy	Dynamic ABR
Segment Size	$\approx$ 2 seconds
scanType	progressive
minBufferTime	PT3.00S
JUJU version	2.4.3
Kubernetes version	1.11/stable
Docker version	18.06.1-ce
NGINX version	1.15.4
Google Chrome version	68.0.3440.106 (64-bit)
Client to 3rd-tier (Cloud) link delay	200ms
Client to 2nd-tier (Fog) link delay	70ms
Client to 1st-tier (Edge) link delay	22ms


DASH Industry Forum
Reference Client 2.9.0
Stream
https://192.168.1.1/akamai/bbb_30fps/bbb_30fps.mpd
Show Options
Stop
Load
768x432 / 1500 kbps / 30 fps
Frame 1550: PTS= 00:00:51.667
00:51
10:34
Video
Audio
Buffer Length : 20.728
Bitrate Downloading : 1884 kbps
Index Downloading : 6
Current Index / Max Index : 6 / 10
Dropped Frames : 0
Latency (min|avg|max) : 0.53 | 0.82 | 1.45
Download (min|avg|max) : 1.44 | 1.63 | 2.01
Ratio (min|avg|max) : 1.99 | 2.46 | 2.78

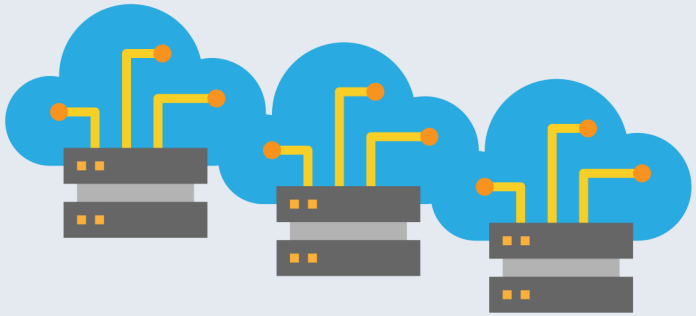

Cloud
Cloud Servers
Internet
Multi-tier Fog/Edge
Core Network
Regional EDGE
Access Network
EDGE Nodes
EDGE
Gateways
Devices
Cyber Physical
layer
B.1
B.2
C.1
C.2
A
B
C
Cache
High resolution
Transcoding
Low resolution
Transcoding
Augmented
Reality
360⁰
Video

Cloud and Edge Computing!

Size

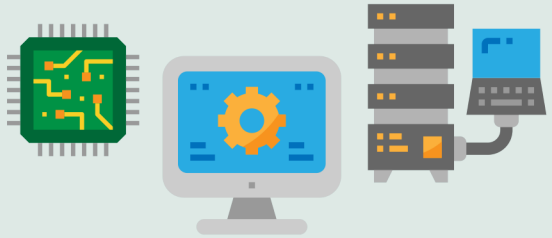
Cloud

Big data processing
Data warehousing
Business logic



Edge

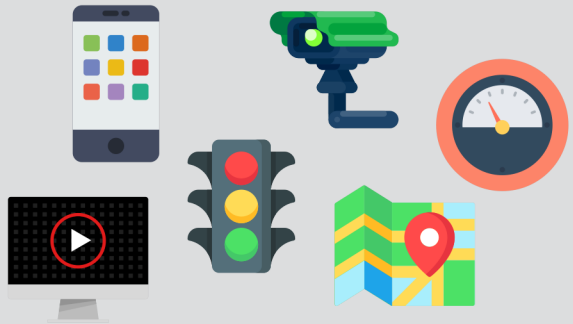
Realtime processing
Analytics
Data caching, filtering



Objects

Heterogeneous devices

Multi-access connectivity



Latency