The video and transport protocol stack for 5G

Brian Stevenson

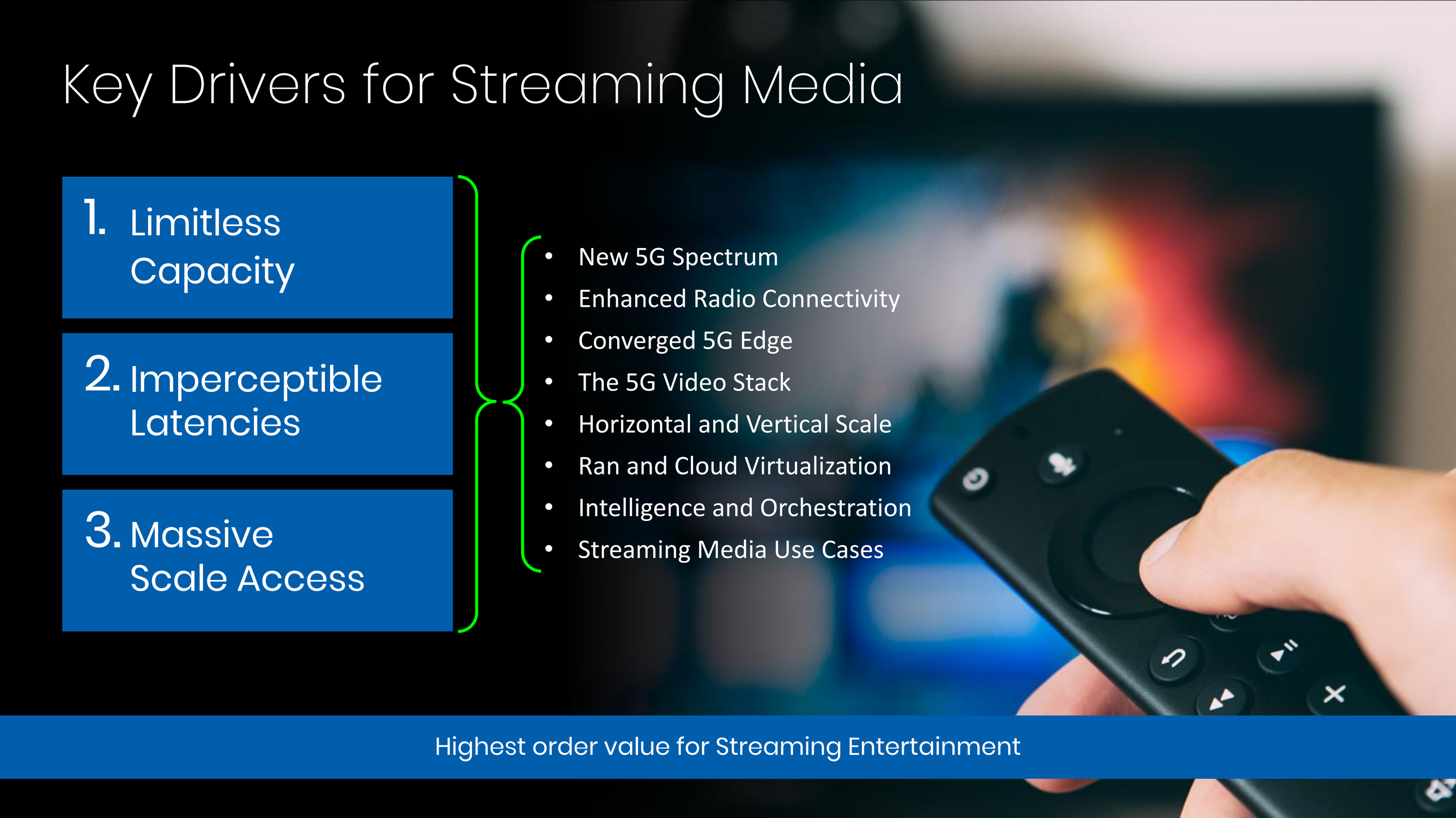
June 17th, 2021

streaming video alliance

Key Drivers for Streaming Media

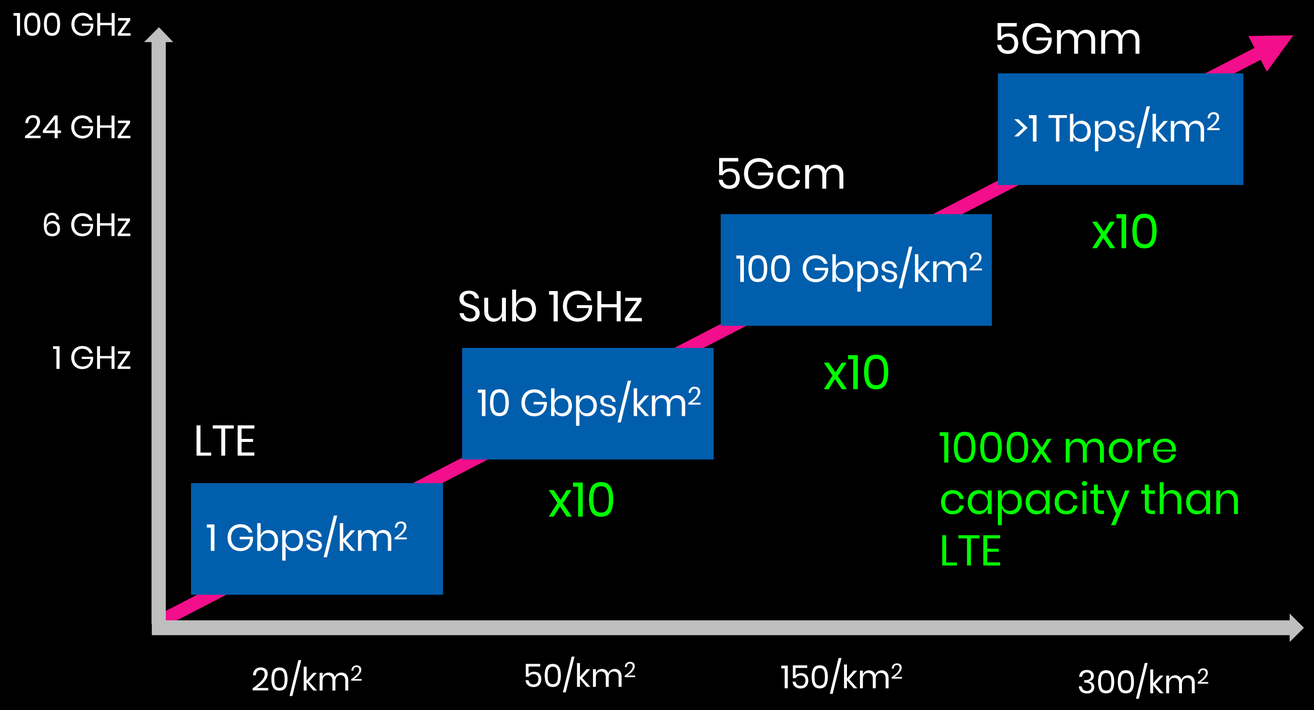
1. Limitless Capacity
2. Imperceptible Latencies
3. Massive Scale Access

- New 5G Spectrum
- Enhanced Radio Connectivity
- Converged 5G Edge
- The 5G Video Stack
- Horizontal and Vertical Scale
- Ran and Cloud Virtualization
- Intelligence and Orchestration
- Streaming Media Use Cases



Highest order value for Streaming Entertainment

Massive Scale Access



New 5G Spectrum

- "Low-band" (sub-1GHz)
  - Currently used for LTE
  - Limited Bandwidth – 100Mpbs shared
  - 20ms RTT
  - Bandwidth is limited/depleted
- Mid-band (sub-6GHz or 5Gcm)
  - Leverages the sub-6GHz spectrum
  - 1Gbps/site (shared)
  - 5-10ms RTT
  - Sweet spot for streaming and XR applications
  - Narrow coverage
  - Limited Object Penetration
- mmWave (> 24GHz)
  - 10Gbps/site (shared)
  - Extreme low latency <4ms RTT
  - Low coverage (250m)
  - Poor Object Penetration

High Frequency = Fast Speeds/Short Distance… Low Frequency = Slow Speeds/Long Distance

Enhanced Radio Connectivity

1. Small Cells

- MM Wave base stations
- Placed @ 250m in urban areas
- Cells form a dense network which provides exceptional connectivity
- Cells provide massive spectral efficiency through frequency reuse
- Infill network bridging the gap between Cellular and WiFi

2. Massive MIMO

- 100 Antenna Ports and dozens of antennas on a single array
- Increases the volume of mobile network users by more than 20.
- A drawback is cross interference due to the density of cellular traffic

3. Beamforming

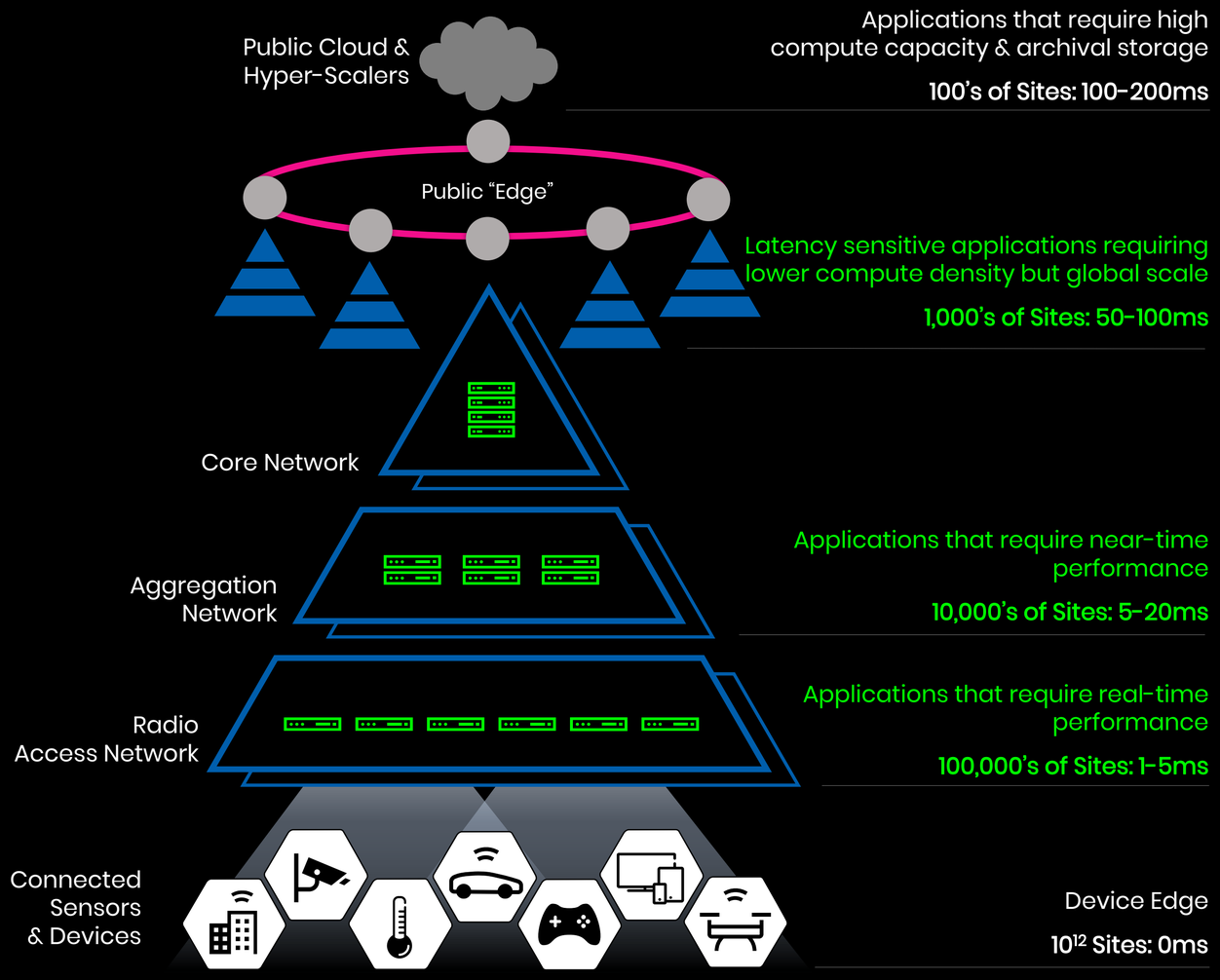
- Packet delivery is carefully orchestrated eliminating interference
- Avoid objects by concentrating the beam only at the user device
- May be bounced of objects to optimize the path

4. Full Duplex

- Transmits and receives at the same time on the same frequency.

Enhanced radio connectivity is the essential enabler of ultra-capacity 5G RAN

A Converged 5G Edge



Applications that require high compute capacity & archival storage

100's of Sites: 100-200ms

Latency sensitive applications requiring lower compute density but global scale

1,000's of Sites: 50-100ms

Applications that require near-time performance

10,000's of Sites: 5-20ms

Applications that require real-time performance

100,000's of Sites: 1-5ms

Device Edge

10^{12} Sites: 0ms

Key functions & applications migrate down into the edge cloud to localize traffic and reduce latency

Low-latency
Video caching and user plane functions close to the access drives interactivity

Massive Capacity
Seamless scale for personalized UHD & xR entertainment

Tera-scale
Optimized connectivity for 1 Billion devices

SW-defined video functions migrate up into the edge cloud for better agility, scale and reduced TCO

Distribution of ultra-small devices achieve real-time performance

Scale, flexibility, and programmability for new distributed streaming entertainment services

The 5G Video Stack

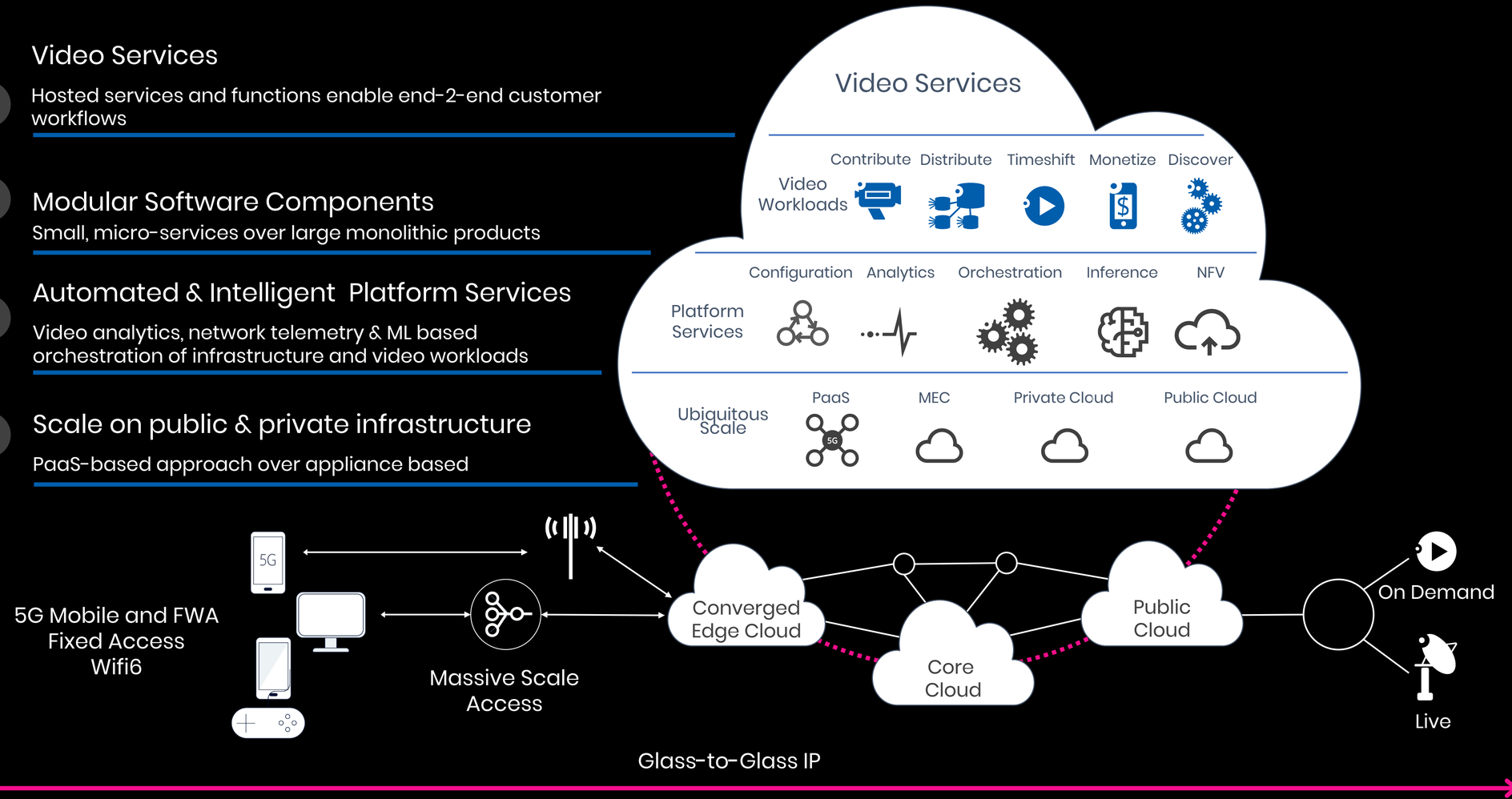
Video Services

1. Hosted services and functions enable end-2-end customer workflows

2. **Modular Software Components**
   Small, micro-services over large monolithic products

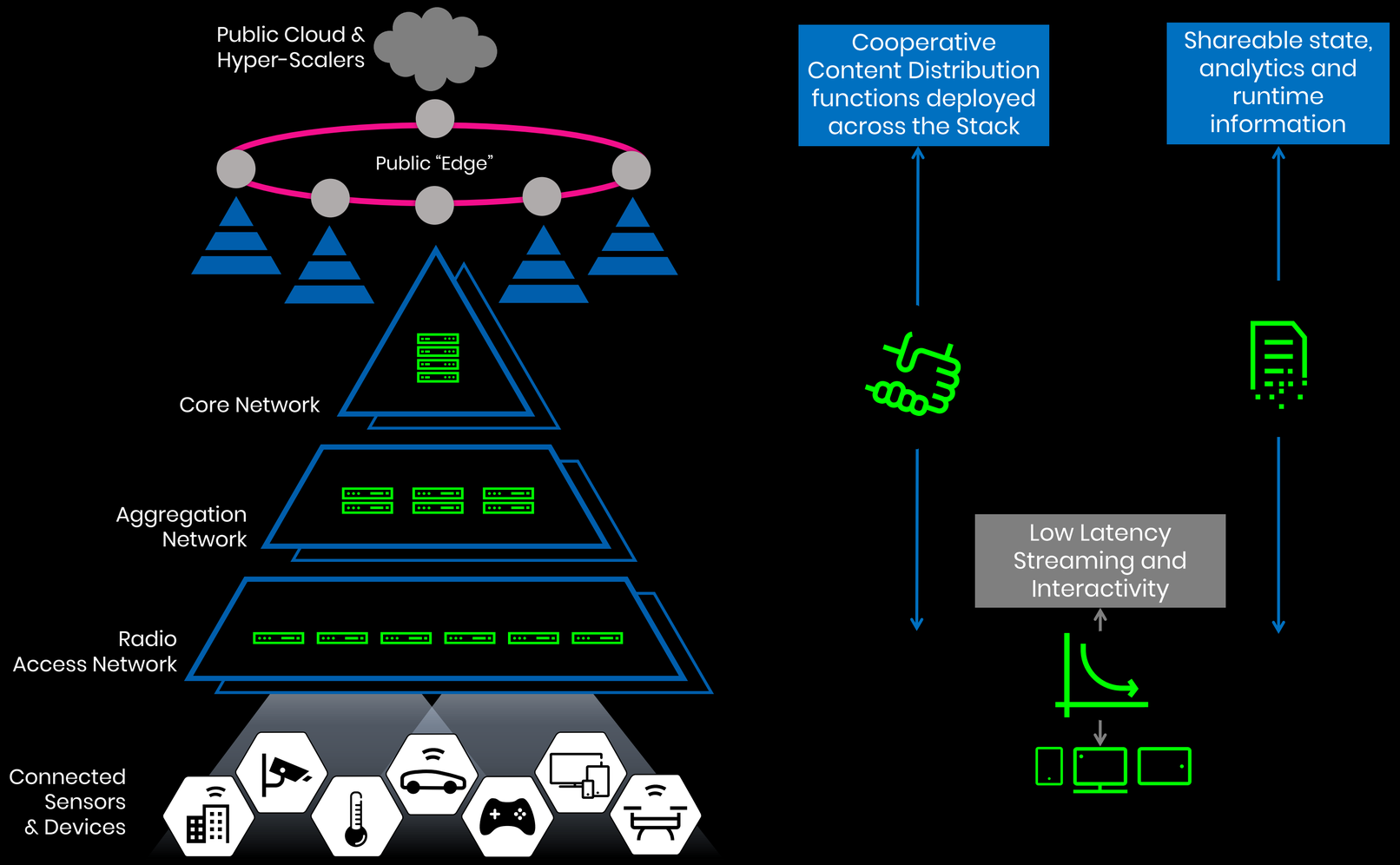
3. **Automated & Intelligent Platform Services**
   Video analytics, network telemetry & ML based orchestration of infrastructure and video workloads

4. **Scale on public & private infrastructure**
   PaaS-based approach over appliance based



Shared compute and common big data infrastructure drives massive efficiencies and lowers TCO

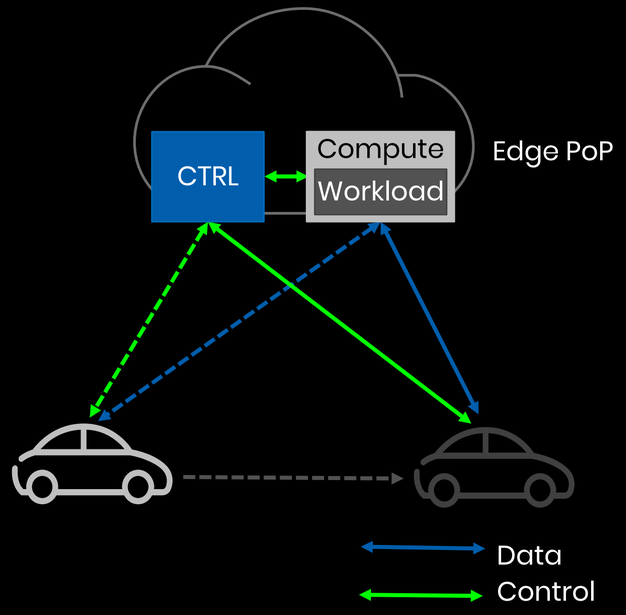
Vertical Scale for Video Applications



- There is no "Right" place to run Streaming applications
- If the edge cloud exceeds its computational capacity the excess workload can be offloaded to a peer or higher tier.
- Allows to scale peak loads by provisioning computational capacity anywhere in the Edge to Cloud Stack.

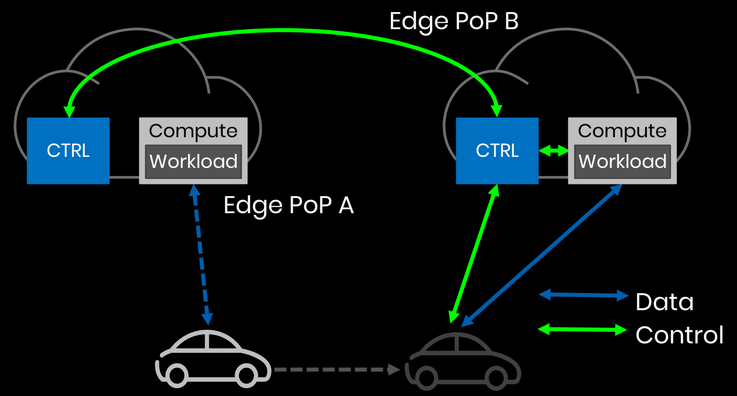
Vertical scale delivered through the "Edge to Cloud" Stack

Horizontal 5G Scale



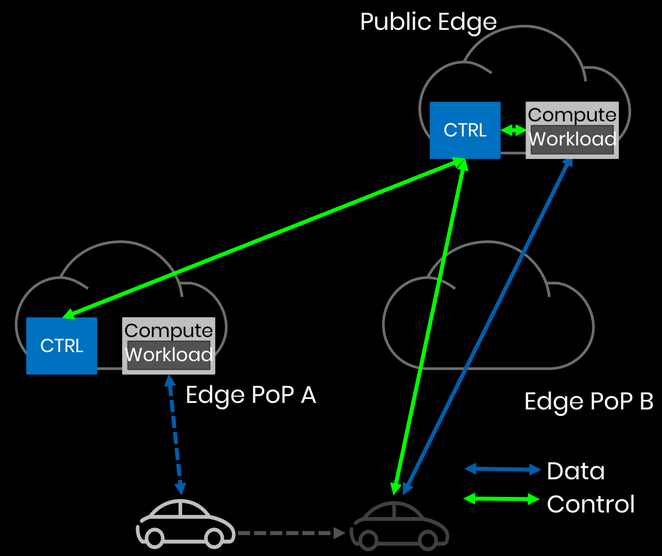
Intra Edge

- **Scenario**
  - Device moves from one eNodeB to another but remains in the coverage of the serving Edge PoP.
- **Workflow**
  - Edge PoP routes traffic to the device via a new eNodeB and tunnel.
  - Assumes there multiple eNodeB's associated with a single Edge PoP.



Inter Edge 1

- Scenario
  - Device moves out of the coverage area of the original Edge PoP and into the coverage area of a new PoP or Network Operator.
- Workflow
  - The originating Edge PoP relocates the application and state, on a new target PoP.
  - The Local Core informs the local Edge instance about device move. Local Edge notifies the remote Edge in new location to move the execution of the application.



Inter Edge 2

- Scenario
  - Device moves out of the coverage area of the original Edge PoP and into the coverage area of a new Network Operator that has no Local Core or no Edge Compute
- Workflow
  - The originating Edge PoP relocates the application and state into the Edge to Cloud Stack
  - Local Core inform Edge instance to move execution to the Public Cloud

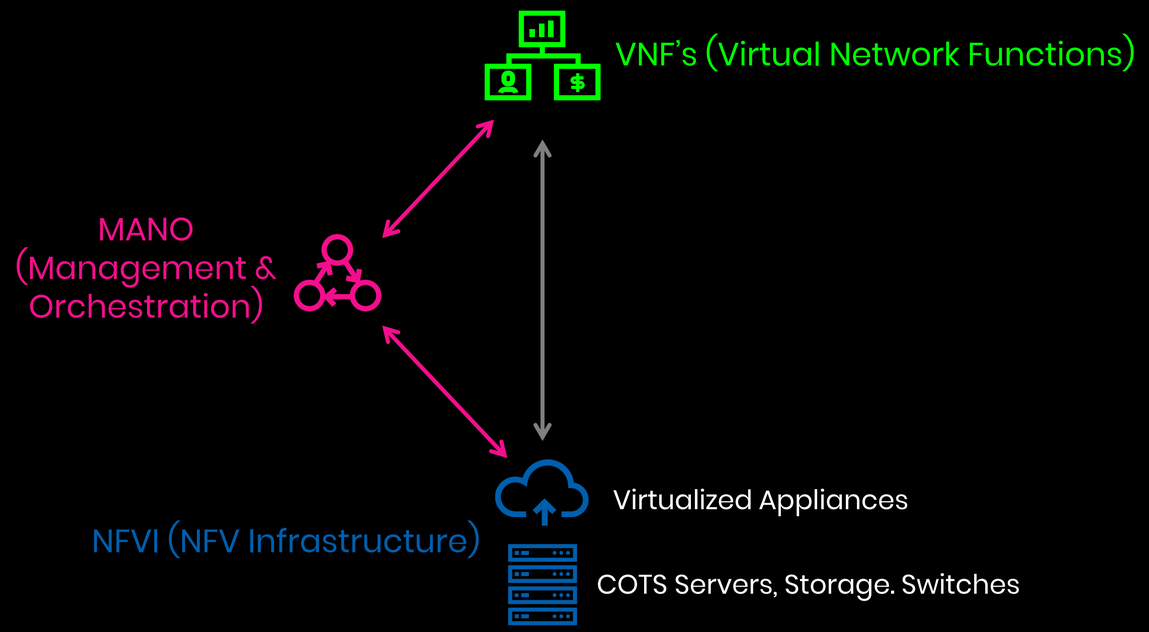
Horizontal scale for seamless Mobile access

Agile Virtual Infrastructure & Functions

5G Network Function Virtualization (NFV)

- NFV Functionality
  - Decouples software from hardware
  - Virtual Network Functions (VNF) run in PaaS or IaaS.
  - Programmable Network OS.
- Self Organizing Networks and Clouds
  - E2E NFV platform and resource optimization
  - Service Orchestration and lifecycle management
  - Flexible distribution, scaling of edge and core functions
- Modular
  - Open source, open API and multi-vendor
  - Cloud native workloads with small capabilities
  - Common “Big Data” infrastructure
- Network Slicing
  - Enables multiple virtual networks to run simultaneously.
  - Guaranteed QoE for Stream and Television Services

NFV Architecture



5G Network and Edge Cloud decomposition into network functions

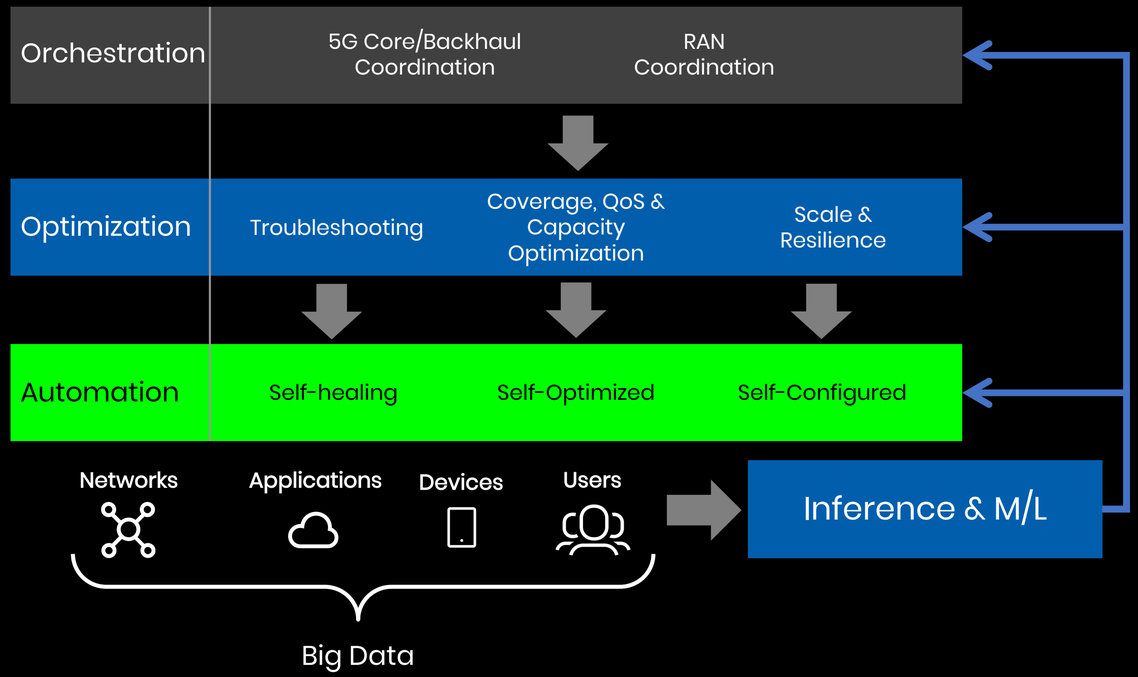
Intelligent and Self Organizing

1. Predictive scaling and resource assignments based on insights and inference

2. Network and application anomaly, fault detection self-healing and optimization drives considerable improvements in QoS

3. Automatic threat detection and immunization mitigates malicious attacks and attack mitigation.

Self Organizing Networks



Increasing network and cloud complexity requires cognitive control and automation

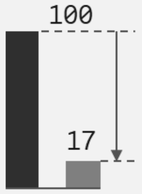
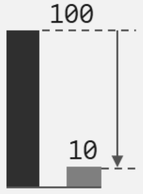
Agile and Orchestrated

Appliance v's Cloud

Time to Launch New Services	
Appliance: **6-7 Months** — Manual Processes	Cloud: **11 Seconds** — Continuous Integration & deployment

# Server/Admin	
Appliance: **<50** — Specialized Servers & Systems	Cloud: **15,000** — Fully Automated Virtual Servers

Ops. vs. Dev/Lean Ops.

$$$ Time to Market	Reduction in Cycle Times
 **83% Reduction** in number of days from code complete to live production	 **90% Reduction** in number of days updating servers and IT environment

Established Infrastructure Vendors

Juniper Networks, Ericsson, Cisco, Nokia, Arista, IBM, F5, Siemens

Web-scale, agile CSP's and Content Owners focus on the right

Hyper-Scalers and Open-Source

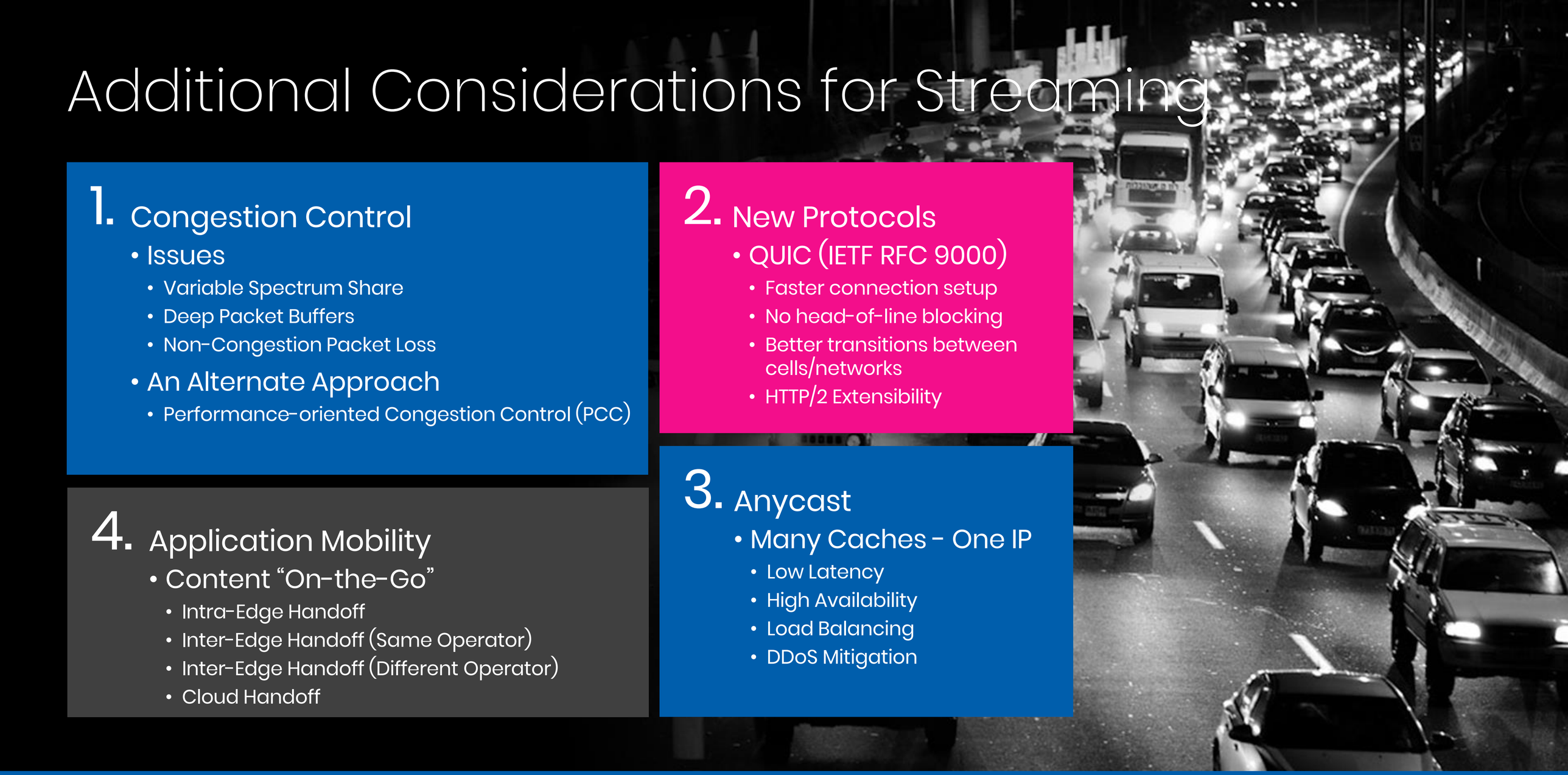
amazon web services, kubernetes, redhat, Azure, docker, OPNFV, openstack, vmware, ONF

By 2022 74% of strategic Infrastructure relationships will have shifted to the right

- Gartner, Cloud Strategy Leadership Report

Shift to cloud architectures requires DevOps & Lean Ops to realize full value potential

Additional Considerations for Streaming



1. Congestion Control

- Issues
  - Variable Spectrum Share
  - Deep Packet Buffers
  - Non-Congestion Packet Loss
- An Alternate Approach
  - Performance-oriented Congestion Control (PCC)

2. New Protocols

- QUIC (IETF RFC 9000)
  - Faster connection setup
  - No head-of-line blocking
  - Better transitions between cells/networks
  - HTTP/2 Extensibility

3. Anycast

- Many Caches - One IP
  - Low Latency
  - High Availability
  - Load Balancing
  - DDoS Mitigation

4. Application Mobility

- Content “On-the-Go”
  - Intra-Edge Handoff
  - Inter-Edge Handoff (Same Operator)
  - Inter-Edge Handoff (Different Operator)
  - Cloud Handoff

OTT subscribers receive HD content less than 40% of the time - Wall Street Journal Report, 2019

M&E Use Cases

1. Enhanced OTT

- The “Killer” use case for 5G?
- Traditional TV, Streaming Services and Extended Reality
- 5G and FWA
- FeMBMS
- Network Slicing
- Content-Aware Workloads

2. Cloud Gaming

- Virtual Gaming Rigs
- Stream Anywhere
- XR and VR Experiences

3. 320MPH Entertainment

- Connected Cars
- Public Transport
- Autonomous Vehicles

4. “In-Venue” Experiences

- Immersive Content
- XR and Mixed Reality
- Insights and Metadata
- Deeper Fan Engagement

5. Content Contribution

- News Gathering, Citizen Journalism, (e-)Sports, etc.
- 5G Uplink – HD, 4K, 360° & VR
- Automated Production & Rights Management

Any surface, any content, any place

Will 5G Challenge the Status Quo

Broadcast and ATSC 3.0 – Will they be…

- Competing Technologies…
  - FeMBMS is the strongest competition for free-to-air broadcast services.
  - 4k and 8k capability
  - H.264, H.265, ISOBMFF - but generally codec and packaging agnostic
  - Provides FTA mobile reception on devices without a registered SIM.
- …or Hybrid Cooperatives
  - ATSC 3.0 primary channel, 5G bespoke content, multi-view, catch-up and rewind
  - Hyper-Monetized targeted advertising
  - Enhanced Functionality – Fast Channel Change, Retransmission
  - Accessibility and Title6 compliance
  - Enhanced Metadata

If deployed correctly, the person likely to benefit most is the consumer

The SVA 5G Technical Brief

- Provide educational resources to improve industry decision making regarding 5G adoption for Streaming
- Examine current 5G and Edge trends, markets and deployments
- Explore 5G and Edge technology focusing on the specific benefits for Streaming
- Define M&E use cases that will explicitly benefit from 5G
- Determine how or if 5G will challenge the Status Quo