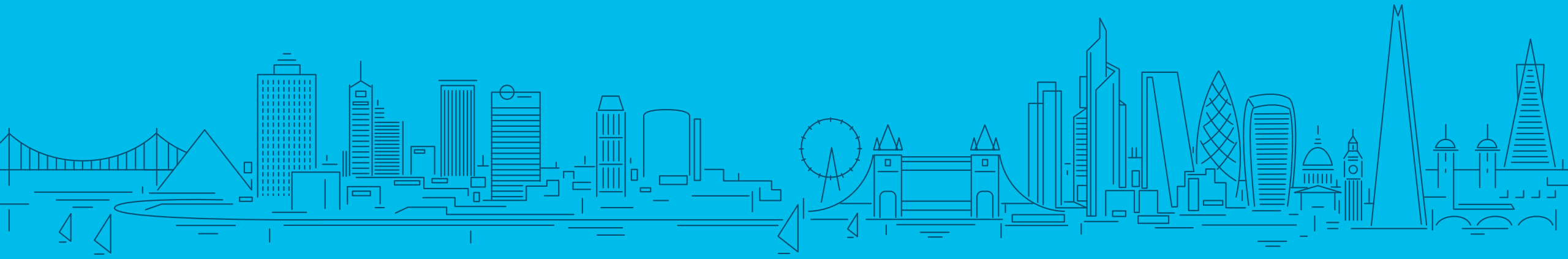


Mitigating RAN congestion with Cisco Ultra Traffic Optimization (CUTO)

Reuben Yap
APJC SP Mobility Architecture

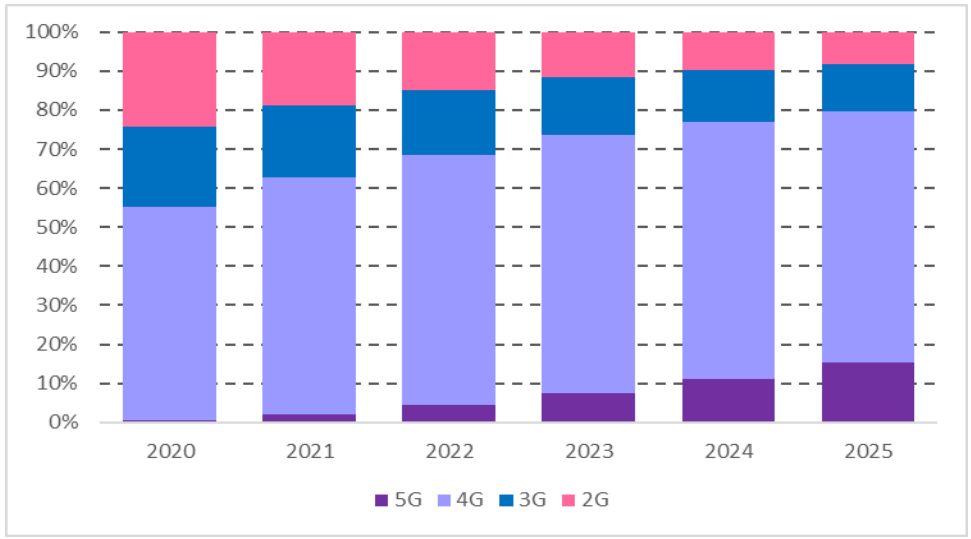
John Burnette
CTO, Opanga Networks

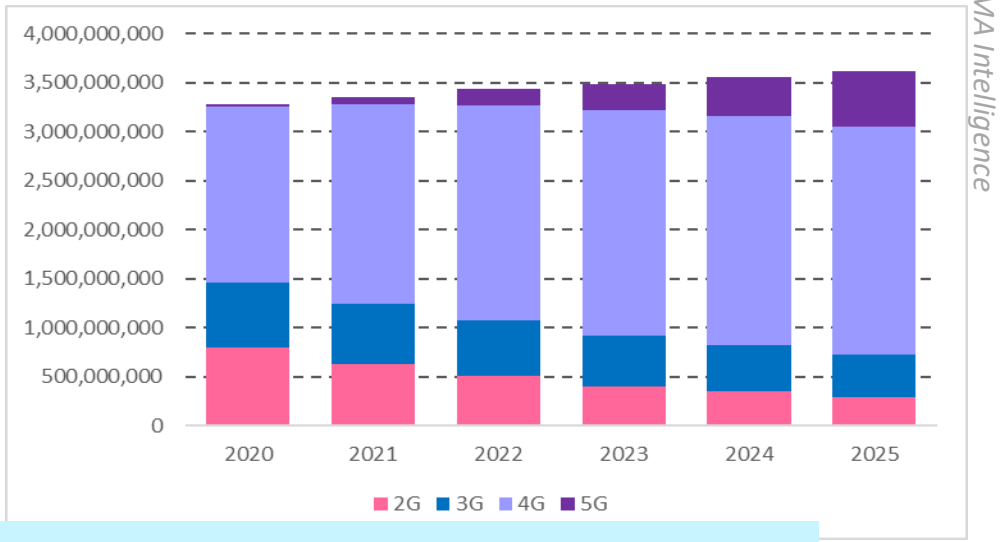
Feb 2022

4G to 5G Evolution

- 5G evolution is in full swing, led by Tier-1 APJC markets such as China, Japan, Korea, Australia.
- 4G continues to dominate in APAC, serving the Mobile Broadband demand and supporting many governments' internet service obligations
- MNOs need to invest in 5G though revenue generation continues to be from existing 4G

Mobile Connections APAC (excl. China)



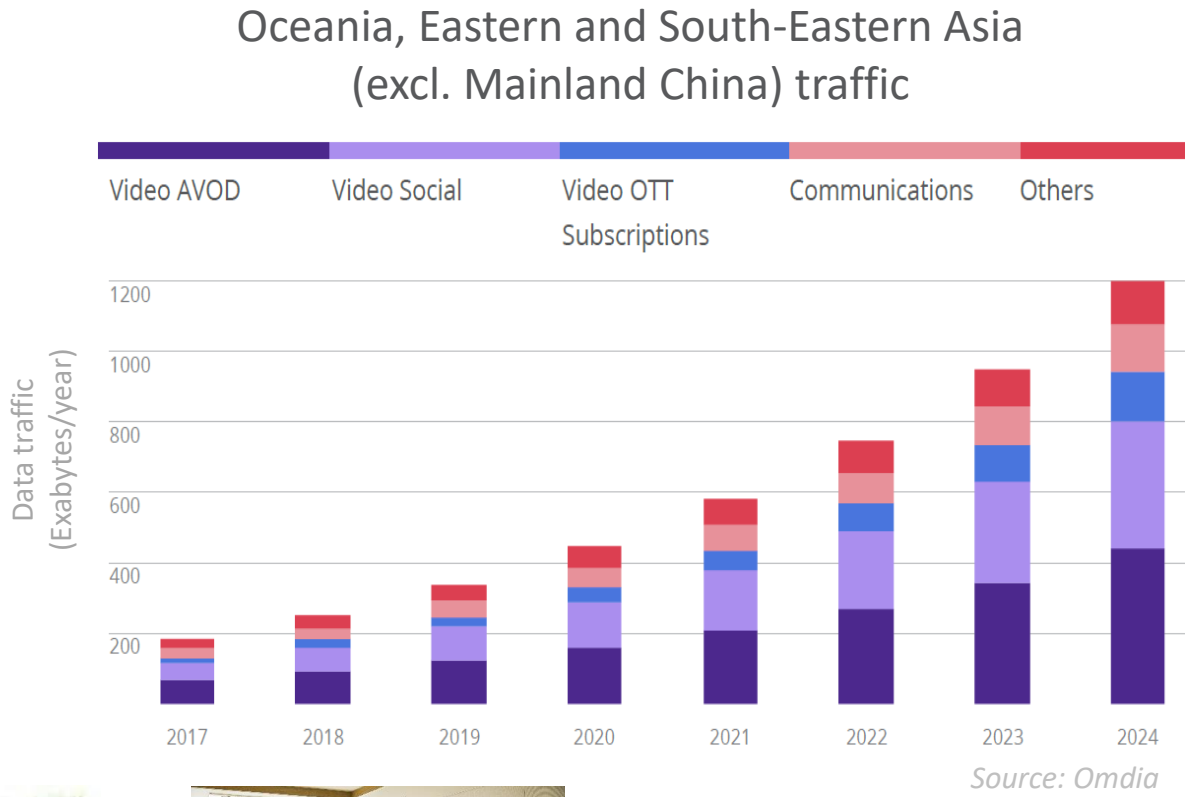


Source: GSMA Intelligence

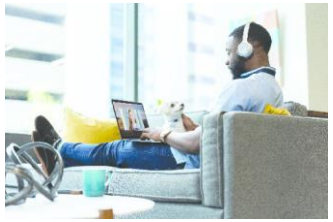
MNOs need to juggle 5G investments while maintaining good service quality on 4G

Video Traffic Continues to Grow

- Future of Work is evolving - connected workforce, remote teleworking, webex ..
- Video based social-media Instagram, TikTok, Facebook, etc
- Proliferation of streaming video/media services – Netflix, Disney+



Source: Omdia



Video Consumption continues to grow on Mobile networks, imperative that MNO's manage customer QoE

Video Drives Network CAPEX/OPEX and Quality

Video sessions consume 20,000x more bandwidth and **destroy eNodeB efficiency** - decreasing spectral and eNodeB capacity.

Video drives CAPEX/OPEX, Quality/Churn, EBITDA – **the industry**!

This is a **Digital Communications** issue most effectively resolved using a **Big Software** approach. eNodeB vendors are not incentivized to solve the issue since it would cannibalize their revenue

What is Cisco's Ultra Traffic Optimization ?

- CUTO is a Machine Learning Algorithm, detects RAN congestion in Real Time and manages flows in the system for a better user experience
- Deployed in /near the core of the network (integrated in Cisco Packet Core or as standalone solution)
- Able to detect congestion in Real Time across all RAN technologies (3G/4G/5G) and is RAN vendor-agnostic
- Can manage all traffic types (TCP, UDP, QUIC, etc) to better utilize existing RAN more efficiently
- Reduces congestion in the RAN network, improves network speeds and optimizes RAN investments

CUTO Demo

CUTO Value Proposition

Time to Market

- Immediate RAN capacity gain without the high complexity, extended lead times, and execution risk associated with traditional RAN hardware densification.
- E.g. Adding spectrum overlay in a dense urban market takes ~2 years vs enabling CUTO for a one carrier virtual spectrum overlay takes ~2hours

Compatible with TCP Optimization

- TCP Optimization technology adjusts the network congestion avoidance parameters of TCP to increase performance
- This is completely complementary to CUTO and can coexist.

All vendors & Technologies

- Works across all RAN vendors (Ericsson, Huawei, Nokia, Samsung)
- Works across all technology generations (3G, 4G, 5G)

Same benefits as adding spectrum

- Enables same network performance and capacity benefits as adding spectrum and cell site hardware to SP networks.
- A 10MHz eNodeB carrying 15Mbps today can now carry 22Mbps
  - Equivalent to adding a 4MHz spectrum band to every cell site

For encrypted & non encrypted traffic

- Optimizes all traffic encrypted or non-encrypted, video and non-video

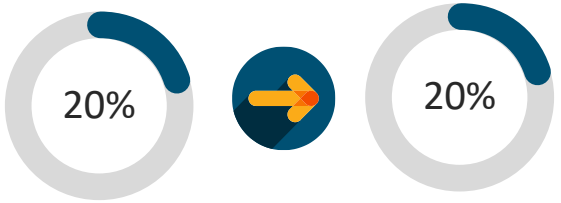
Maintains high-definition video experience

- No change to video start time, stall rate, and resolution.

Packet Core or Standalone

- Can be enabled in Cisco Packet Core with negligible CPU impact or as a standalone solution

Cell Congestion Drives Consumer Experience and CAPEX Requirements

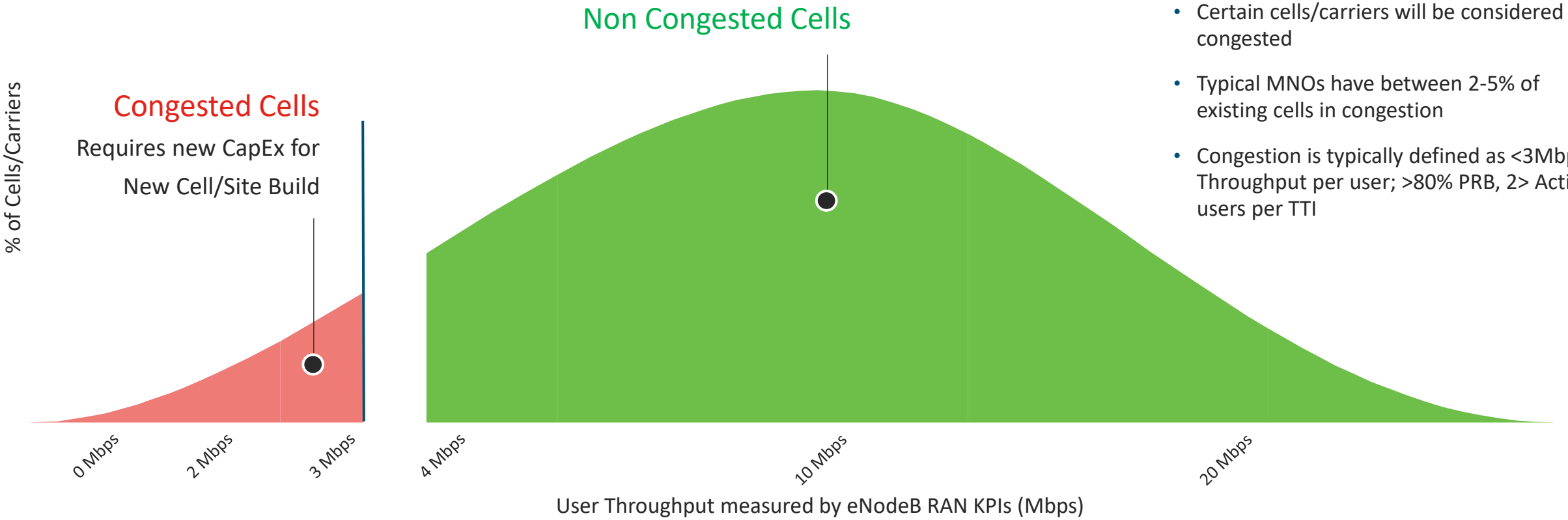


Avg. User Throughput Increase during RAN Congestion

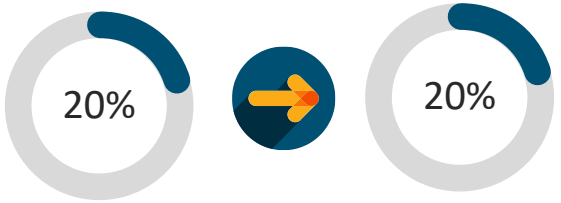
RAN CapEx/OpEx Reduction

When CUTO is OFF

- Certain cells/carriers will be considered congested
- Typical MNOs have between 2-5% of existing cells in congestion
- Congestion is typically defined as <3Mbps Throughput per user; >80% PRB, 2> Active users per TTI

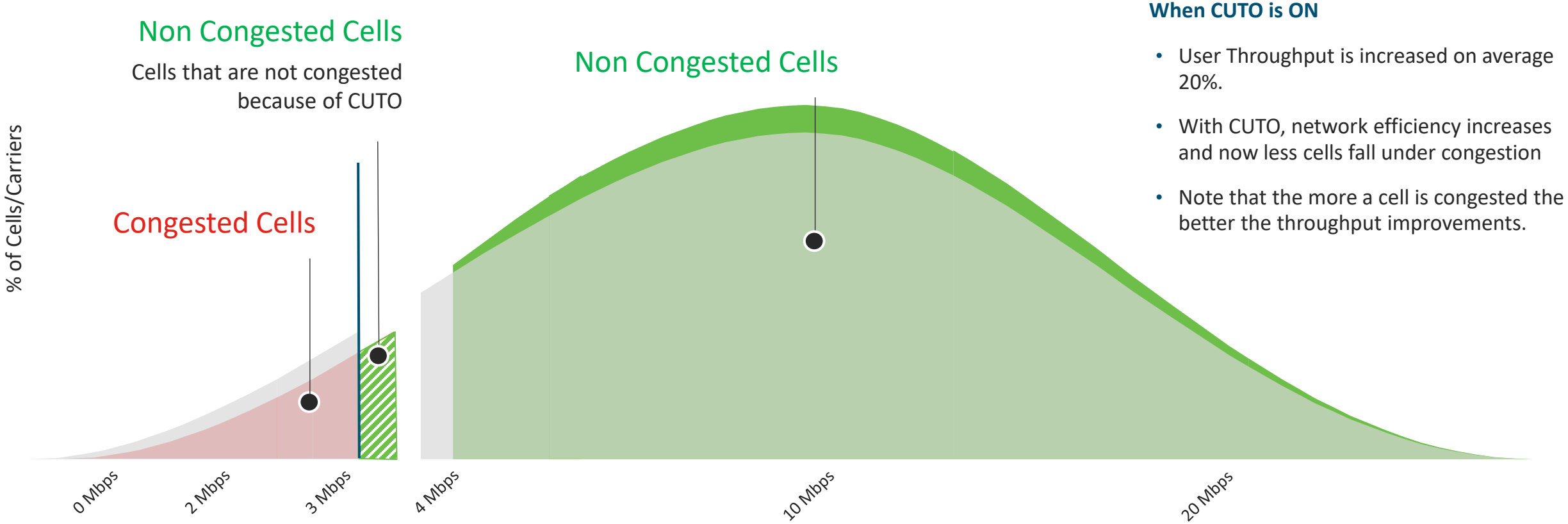


Throughput Increase = Faster Network, Less Congested Cells



Avg. User Throughput Increase during RAN Congestion

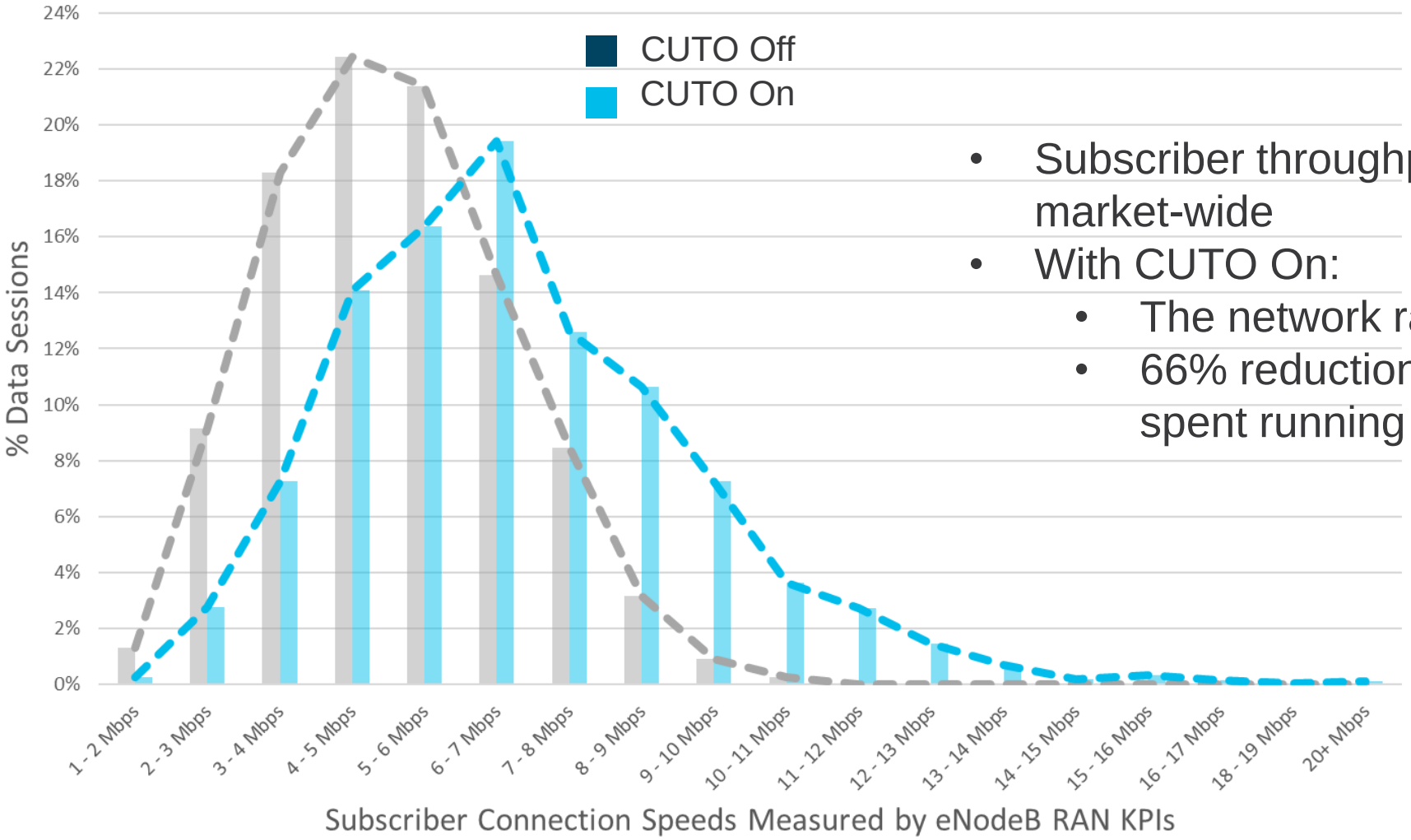
RAN CapEx/OpEx Reduction



When CUTO is ON

- User Throughput is increased on average 20%.
- With CUTO, network efficiency increases and now less cells fall under congestion
- Note that the more a cell is congested the better the throughput improvements.

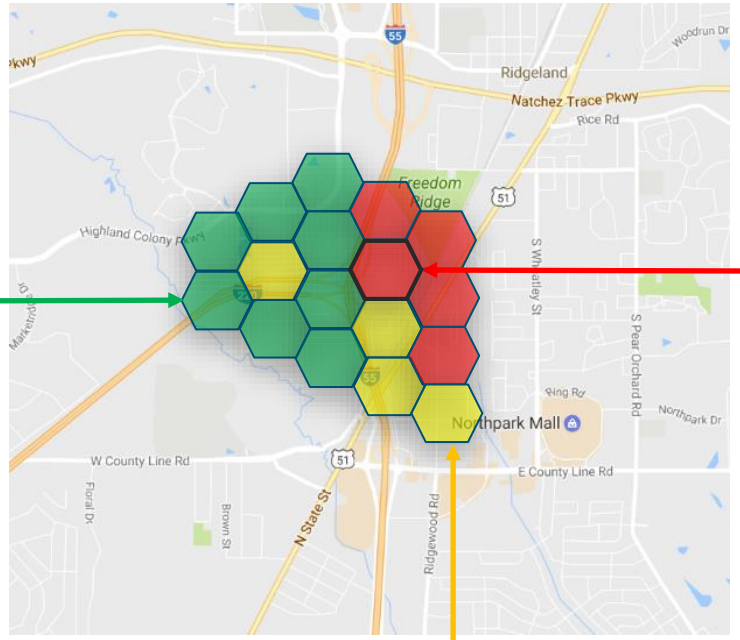
Measured Network Performance Shift



- Subscriber throughput measured market-wide
- With CUTO On:
  - The network ran 34% faster
  - 66% reduction in time the network spent running below 4Mbps

Cellular Network Congestion Mapping
Where is CUTO Providing Gain?

Low Network Resource Contention
- Very Little Concurrent Elephant Flows
- Essentially No Gain – Opportunistic Only
- Insurance if it becomes congested



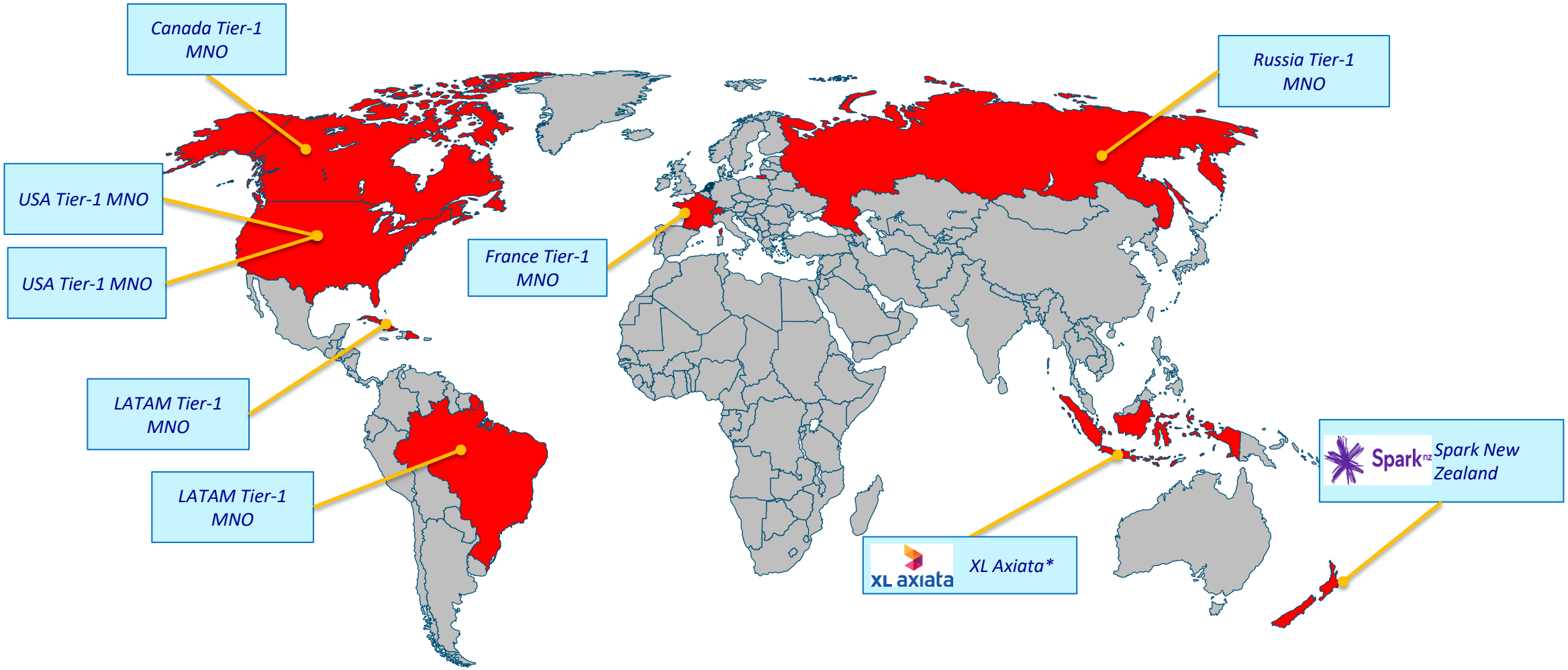
Heavy Network Resource Contention
- High Concurrent large Flows
- Substantial Cell Site Capacity Improvement
- Substantial Connection Speed Improvement
- Substantial Latency Reduction

Moderate Network Resource Contention
- Several Concurrent large Flows
- Material Cell Site Capacity Improvement
- Material Connection Speed Improvement
- Material Latency Reduction

CUTO Deployment Results

Operator	Predominant Service Plan	Traffic breakdown	Throughput Improvement	
			eNodeB RAN KPIs	Ookla Speedtest
US Operator 1	Unlimited Plans	3% of interactions use 68% of traffic volume	Not Measured	+33%
US Operator 2	Unlimited Plans	2% of interactions use 48% of traffic volume	+24%	+29%
US Operator 3	Unlimited Plans	3% of interactions use 54% of traffic volume	+24%	+20%
EU Operator 1	Unlimited Plans Aggressive on Video	3% of interactions use 72% of traffic volume	+34%	+33%
EU Operator 2	Unlimited Plans Aggressive on Video	4% of interactions use 70% of traffic volume	+30%	+44%
EU Operator 3	Limited Plans	3% of interactions use 54% of traffic volume	+13%	+30%
LATAM Operator 1	Limited Plans	1% of interactions use 30% of traffic volume *QUIC Optimization Off	+14%	+25%
LATAM Operator 2	Limited Plans	2% of interactions use 40% of traffic volume	+19%	+22%
LATAM Operator 3	Limited Plans	2% of interactions use 40% of traffic volume	+20%	N/A
APJC Operator 1	Limited Plans	3% of interactions use 60% of traffic volume	+19%	N/A
APJC Operator 2	Limited Plans	4% of interactions use 58% of traffic volume	+17%	N/A

Cisco CUTO Reference Wins



* Non-GW integrated CUTO (Standalone)

Spark press release - https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=2057039

XL press release - https://news-blogs.cisco.com/apjc/2022/01/26/11579/

CUTO References

The Newsroom
Cisco's Technology News Site

Spark

Spark New Zealand Enhances Rugby World Cup Mobile Experience with Cisco Traffic Optimization Software

"The CUTO solution developed by Cisco on their mobile packet core has provided immediate gains in radio network capacity and performance," said Campbell Fraser, Tribe Leader of Technology, Spark. "We deployed the solution nationwide prior to the Rugby World Cup when the streaming of live games created peaks of network demand. We wanted to maximize the user experience for mobile customers, but also ensure critical services - like voice - were protected."

Source: https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=2057039

XL axiata

XL Axiata Boosts Mobile Network User Experience with Cisco CUTO

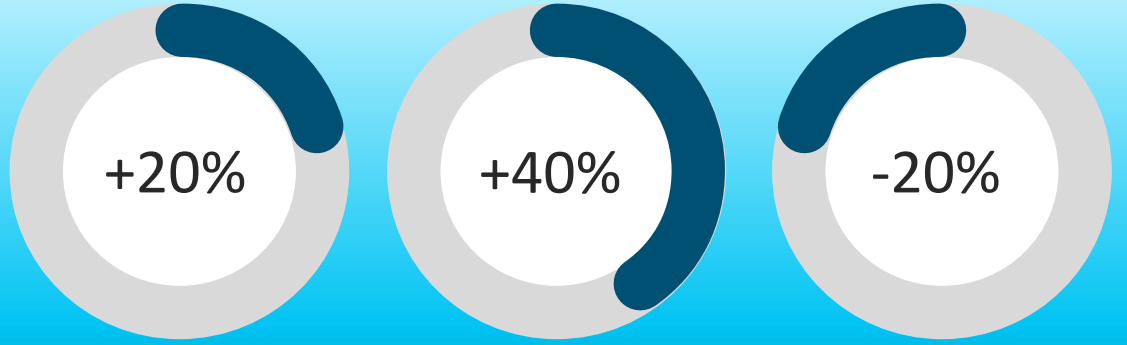
Press Release

"CUTO helps increase the efficiency of network capacity so that our cellular network infrastructure becomes even better. With CUTO, XL Axiata can provide high-quality data services to customers, wherever they are," said Gede Darmayusa, CTO XL Axiata

Source: https://news-blogs.cisco.com/apic/2022/01/26/11579/

CUTO Summary

Provides significant Gains to alleviate RAN Congestion



Provides fast Time-to-Market with flexibility to work across 3G/4G/5G and RAN vendors