White Paper

Cisco Ultra Traffic Optimization: How Mobile Network Operators Can Eat the Network Capacity Augmentation Cake and Have It Too!

Sponsored by: Cisco

Rajesh Ghai
April 2020

EXECUTIVE SUMMARY

This white paper introduces Cisco Ultra Traffic Optimization (CUTO) to the audience, gives readers a deeper understanding of the technology, relays its advantages, and the impact it will have on mobile networking. These takeaways are supported by case studies of customers, such as Spark New Zealand, that have successfully implemented CUTO and experienced significant operational and economic benefits.

Streaming video services, user-generated content, video collaboration platforms, and video games continue to increase in popularity and drive a surge in mobile network traffic around the globe. As the IoT trend takes hold in the next phase of enterprise digital transformation, new video data from video surveillance systems and drones will further increase the amount of video network traffic that mobile network operators (MNOs) will need to deliver on their cellular infrastructure. Today, mobile network traffic spikes during certain hours — for instance, when most consumers indulge in entertainment. Moreover, as the quality of video becomes richer, the size of individual video streams will increase over the foreseeable future.

Consequently, during hours of peak usage and during spikes in activity, the mobile network is often congested — it starts reaching its capacity limits. The congestion occurs mainly because a few video "elephant flows" consume cell site resources to the detriment of nonvideo data flows. This network congestion manifests itself in poor quality of experience for the mobile subscriber, especially for those users who are using the network for nonvideo purposes.

MNOs have a choice to make. Invest in new cellular capacity despite historically low return on capital or not invest and see a degradation in the quality of customer experience. Allowing the degradation of customer experience could mean loss of market share. However, traditional cellular capacity augmentation alternatives such as new spectrum, new cells (macro/small), and SON are capital intensive and take a long time to deploy. MNOs need solutions that can improve the ROI on existing investment, help avoid new capital investment, and maintain the quality of customer experience.

Cisco Ultra Traffic Optimization is an AI-based radio access network (RAN) optimization technology designed to increase cell site capacity by increasing mobile subscriber connection speeds in congested cells. It is a pure software solution that is vendor and technology agnostic. CUTO offers immediate cellular network capacity relief without the high complexity, extended lead times, and execution risk associated with traditional RAN hardware- or software-based network densification

solutions. IDC believes that through its advanced optimization, CUTO can help deliver a better return from RAN and mobile network investments.

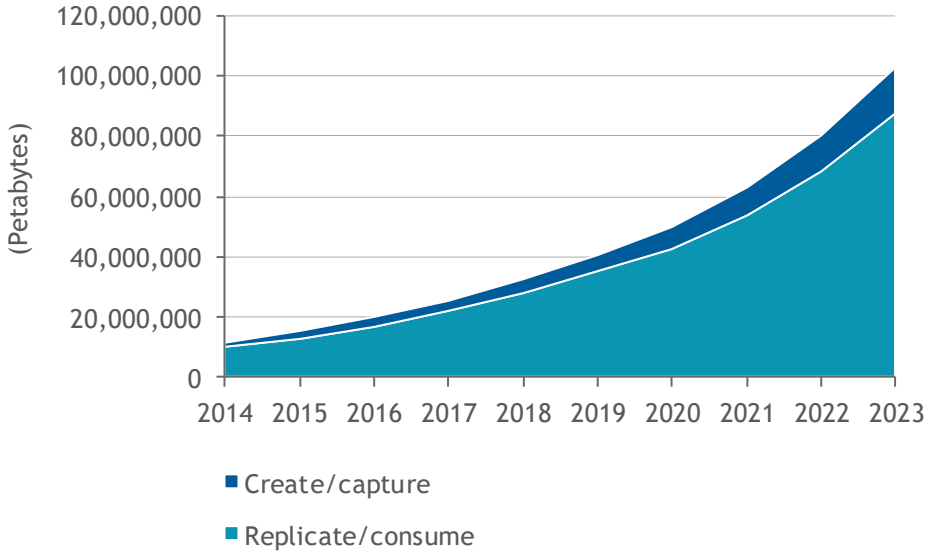
Situational Analysis

Data Continues to Explode

IDC has been sizing and forecasting data creation, capture, and replication for over a decade and the firm calls this data universe the Global DataSphere. According to IDC's 2019 forecast, the Global DataSphere is expected to grow to 102.6ZB by 2023, representing a CAGR of 25.8% for 2018-2023 (see Figure 1).

FIGURE 1

IDC's Global DataSphere Forecast, 2014-2023



Source: IDC, 2019

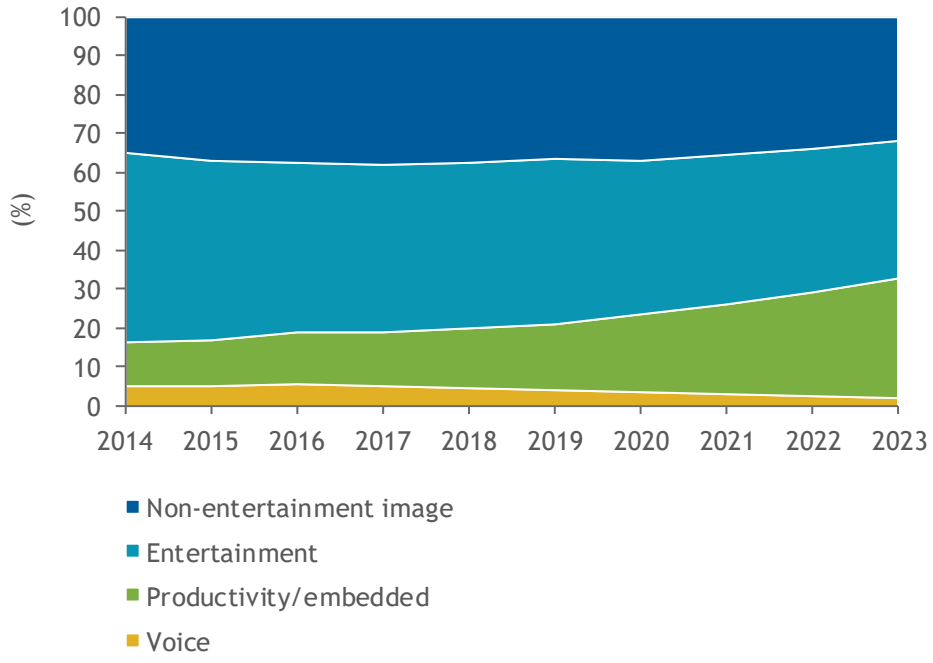
The largest type of data within the Global DataSphere is entertainment video. As might be expected, consumers drive most of the entertainment-related data creation and consumption. This category includes data created for the purpose of creating, producing, distributing, and consuming entertainment content. It includes data from devices such as digital television/radio and consumer mobile devices such as cell phones and tablets. It includes streaming video services, such as Netflix and Hulu; movies; YouTube; and other user-created videos and images, video games, and music. Another key type of video traffic that has steadily grown on mobile networks and has come to the limelight recently is traffic generated by video collaboration platforms such as Webex and Zoom.

In regard to streaming video services, user-generated content, collaboration platforms, and video games continue to increase in popularity; it is not surprising that this not only is the largest component of data growth in the Global DataSphere but is expected to continue to be the largest component over the next decade. As the IoT trend takes hold in the next phase of digital transformation in the enterprise, new video data from video surveillance systems and drones will further increase the

amount of video data being generated. IDC forecasts that in the coming years, video data from these new mobile devices will be a major driver of data and bandwidth growth (see Figure 2).

FIGURE 2

IDC's Global DataSphere: Consumer Entertainment Video — Largest Component of Data Growth



Source: IDC, 2019

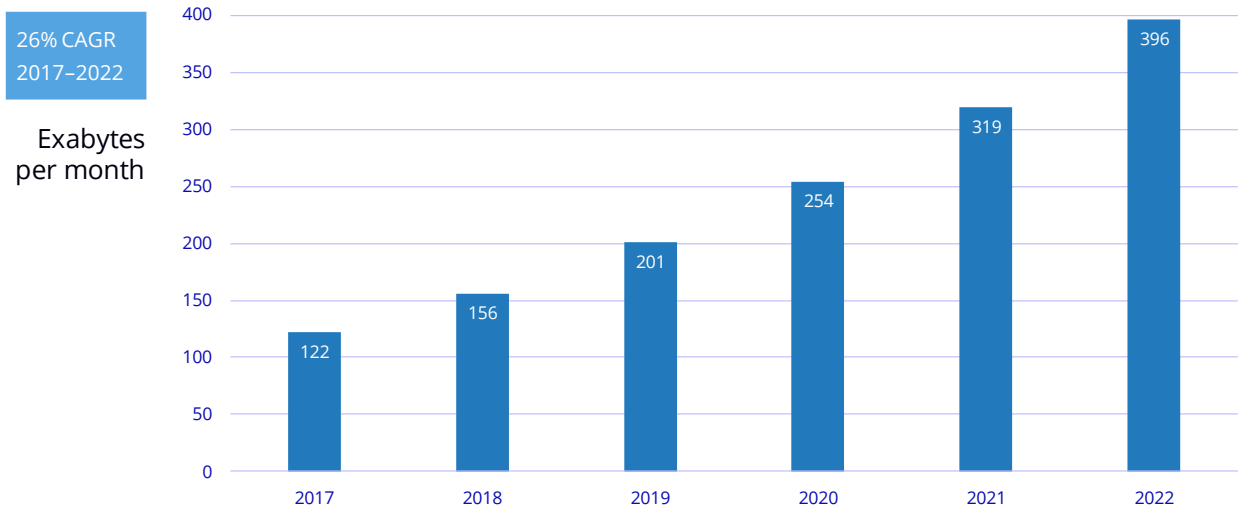
Bandwidth Demand Continues to Surge

Growth in the global DataSphere driven mainly by consumer video entertainment is not, surprisingly, driving a surge in demand for bandwidth from communications service provider (SP) networks.

The Cisco Visual Networking Index (VNI) forecast projects global IP traffic to nearly triple from 2017 to 2022. Overall IP traffic is expected to grow to 396EB/month by 2022, up from 122EB/month in 2017, a CAGR of 26% (see Figure 3). Importantly from a communications SP perspective, these estimates provided by the VNI study appear to be conservative. The growth represented in the 2018 study reflects a slight increase in expectations over the 2017 forecast, which projected a CAGR of 24% from 2016 to 2021, driven by an increase in the growing share of mobile traffic as a percentage of the total IP traffic (see Figure 3).

FIGURE 3

IP Traffic, 2017–2022



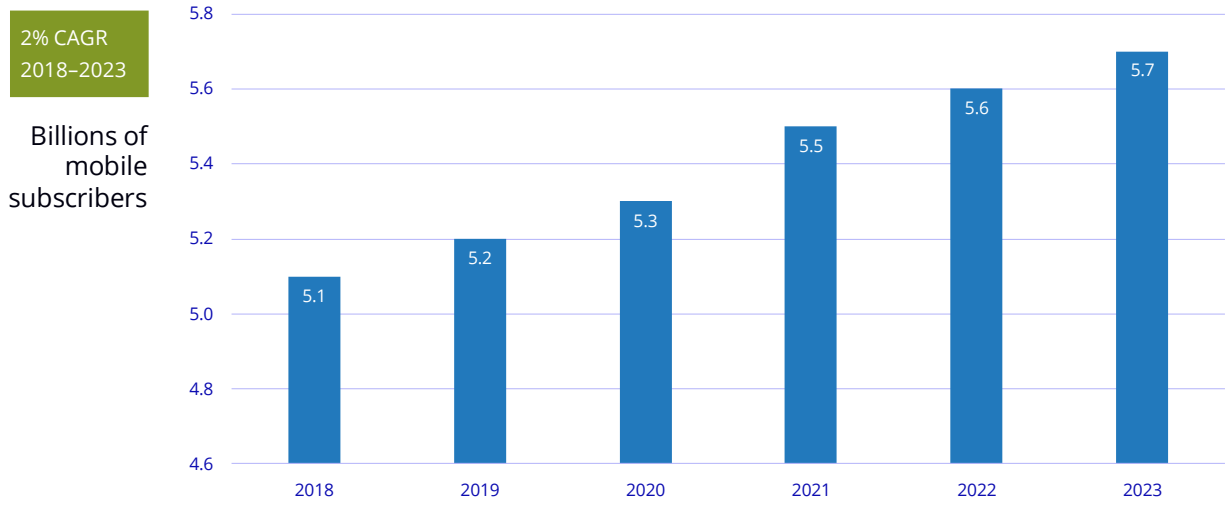
Source: Cisco's VNI Global IP Traffic Forecast, 2018

Mobility Is on the Rise and Dictating How Video Gets Consumed

Globally, the total number of mobile subscribers (those subscribing to a cellular service) is expected to grow from 5.1 billion in 2018 to 5.7 billion by 2023 at a CAGR of 2%, according to Cisco's 2020 VNI forecast. More people around the globe are likely to access their data, apps and, most importantly, video entertainment on their mobile devices than ever before. The mobile device is likely to become the main device for content consumption in the future (see Figure 4).

FIGURE 4

Consumer Mobile Subscribers, 2018–2023: Continued Growth as Mobility Takes Hold



Source: Cisco's VNI Report, 2020

The conclusion that can be derived by considering all these trends together is that consumer mobile broadband traffic is likely to continue growing significantly over the best part of the next decade and the primary driver of this bandwidth demand will be entertainment video. The other key insight that can be drawn from this data is that since most of the network bandwidth is likely to be entertainment video, it is likely to spike during certain hours when most consumers are likely to indulge in entertainment. Moreover, as the quality of video continues to become richer, the size of individual video streams is likely to continue to increase over the foreseeable future. These large individual video streams are also referred to as "elephant flows."

These trends have implications for cellular network capacity planning. Mobile network operators will not only have to invest in network capacity to address the overall surge in network traffic but also need to invest keeping in mind the expected spike in traffic during peak entertainment hours as well as the continuing increase in the size of individual video streams.

5G Will Further Accelerate Mobile Video Traffic Growth

5G networks are expected to increase peak mobile data speeds from current 4G LTE speeds of 100Mb–1Gb to speeds of 10Gb. More importantly, the business case for 5G rests on the ability to utilize the 5G network infrastructure for new incremental use cases beyond enhanced mobile broadband (eMBB) such as IoT, smart factories, cloud gaming, and AR/VR that leverage the low-latency and connections capabilities of 5G infrastructure. A lot of these new use cases such as AR/VR and cloud gaming will require the transmission of rich high-definition video. As more bandwidth becomes available on mobile networks, streaming video entertainment producers will rush to deliver even more high-definition video to mobile devices. All these trends will likely further accelerate mobile video traffic growth even above what is expected at present.

Network Congestion and KPI Deterioration in MNO Networks

Mobile video traffic over the past few years has grown faster than anyone expected. The result of this is that during hours of peak usage and during spikes in activity, the mobile network is often congested. In other words, the cellular network starts reaching its capacity limits. The congestion occurs mainly because a few video "elephant flows" essentially consume cell site resources to the detriment of nonvideo data flows. This network congestion manifests itself in poor quality of experience for the mobile subscriber, especially those users who are using the network for nonvideo purposes.

A key performance indicator (KPI) that mobile network operators consider when thinking of cellular network capacity at any location is cell site capacity in terms of available bandwidth per user, or user equipment (UE) throughput. Cell site capacity is defined by the amount of traffic that can be delivered across the radio channel while still providing users a target connection speed per user. Another metric that MNOs consider while doing cell site capacity planning is scheduler latency.

Another key metric of cellular network capacity is network density. One way to measure the density of a mobile network is by dividing the bandwidth per user by the geographic area served. In other words, what bit rate can the cell site provide to how many users in its vicinity? This network density is an important metric used to evaluate network health and effectiveness. Network technologies aim to improve network density in one way or another through hardware, software, or both (discussed in the Mobile Networks Demand More Capacity section). When network density cannot be increased by software and hardware, MNOs typically have no choice but to invest capital in either new cell sites or spectrum.

When network congestion affects a network, it results in deterioration of network density, UE throughput, and scheduler latency.

MNOs Globally Remain Capital Constrained and Challenged to Improve Network Monetization

Communications service provider industry financial data suggests that communications SPs remain severely capital constrained despite several attractive opportunities to invest capital. According to the IDC Global Telecom Indicators database, carriers globally have seen their revenue, operating income, and capex trend flat to down over the past five years (see Table 1).

TABLE 1

Global Telecom Financials, 2014-2019

	2014	2015	2016	2017	2018	2019
Revenue ($T)	1.75	1.62	1.62	1.63	1.64	1.62
Operating expense ($T)	1.35	1.26	1.25	1.26	1.29	1.21
Operating income ($B)	304	282	277	270	282	283
Capital expense ($B)	333	331	310	301	295	291

Source: IDC's Global Telecom Indicators database, 2020

A few observations from this data:

- Carrier revenue has declined from levels seen in 2014 and held flat over the past four years. This trend has occurred despite communications SPs adding more users and more services globally across their networks. Network monetization has been a significant challenge, as reflected in declining ARPU and declining revenue.
- Carrier capex is constrained by a carrier's ability to generate operating income. Carriers are investing almost as much as they are earning in terms of operating income.
- The key to modernizing their networks and providing improved customer experiences and new services lies in carriers' ability to rein in operating expense and hence drive operating income – that can then be invested in new services.
- Improving service provider and internet economics is predicated on service providers optimizing both capex and operating expense.
- Solutions that can enable mobile network operators to improve the ROI on existing investment is the need of the hour to improve service provider economics.

Fierce Competition Makes Compromising Customer Experience Unacceptable

As capital for new investment remains scarce in the face of limited ability to monetize any new investment, mobile network operators have a choice to make. Invest despite the low return on capital or not invest and see a degradation in the quality of customer experience. Not investing in new cell site capacity has historically meant that mobile network operators have had to resort to throttling traffic for their heavy users. This results in a double whammy of extended peak periods and reduced customer experience for more subscribers.

However, competition among mobile network operators, especially in the larger markets around the globe, remains fierce. Allowing the degradation of customer experience could mean loss of market share. IDC strongly believes mobile network operators are in a dire need of solutions that can enable them to improve the ROI on existing investment – solutions that can help them avoid new capital investment but at the same time maintain the quality of customer experience.

Mobile Networks Demand More Capacity

As mobile devices proliferate and as more video is being delivered to mobile devices over cellular networks, the demand for more bandwidth on cellular networks continues to surge. However, network capacity within any mobile network operators' cellular network is limited by the specific capacity of cell sites serving a specific geographic area.

To increase bandwidth in a cellular network, the constraint imposed by finite cell site capacity must be alleviated.

As previously mentioned, the key metric of cellular network capacity is network density. One way to measure the density of a mobile network is by dividing the bandwidth per user by the geographic area served. In other words, what bit rate can the cell site provide to users in its vicinity? This metric is also used to evaluate network health and effectiveness. Improving cellular network capacity comes down to improving cell site network density.

Traditional approaches to network densification are expensive and capital intensive. However, since cellular network capacity drives customer experience, MNOs typically have no choice but to invest

capital to improve network density. The traditional alternatives available to MNOs to increase network capacity include:

- Licensing of new spectrum
- Buildout of new radio — small cells or macro cells
- RAN network tuning software such as SON

Besides being capital intensive, these capacity enhancement options are all very time consuming and disruptive to a brownfield network. MNOs seek a new, less capital-intensive nondisruptive solution to the growing problem of constant new cell site capacity enhancements.

Introducing Cisco Ultra Traffic Optimization

What Is CUTO?

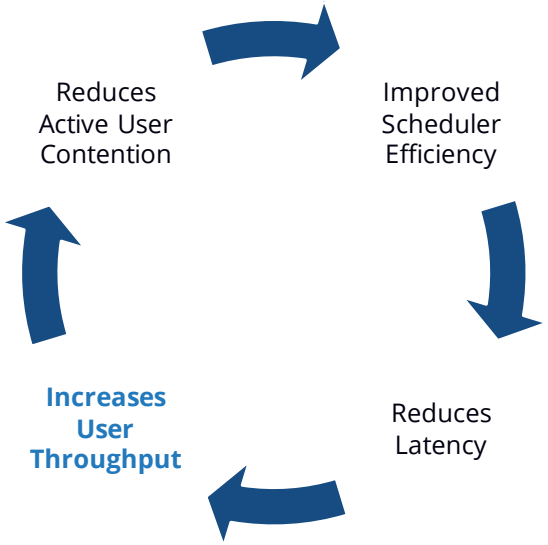
Cisco Ultra Traffic Optimization is a radio access network optimization technology designed to increase cell site capacity by increasing mobile subscriber connection speeds in congested cells. It is fully radio network aware, allowing management on a per-eNodeB/eNodeG cell basis. It is a pure software solution (comprising machine learning algorithms) that integrates with standard Mobile Packet Core (P-GW) including DPI functions such as Application Detection and Control. Using this solution in conjunction with the packet core allows mobile operators to define optimization policies that are based on the traffic and application type as well as APN, QCI, and other common traffic delineations.

Large traffic flows, such as video, during periods of network congestion saturate radio resources and swamp the eNodeB scheduler, affecting key metrics such as UE throughput and scheduler latency. CUTO employs machine learning algorithms to detect large traffic flows (such as video) in the network and optimize the delivery of those flows to mitigate network congestion without altering user quality. In other words, by employing software intelligence at the network core, CUTO mitigates the overwhelming impact video has on the RAN.

CUTO achieves its subscriber connection speed gains by shaping video traffic. It monitors and profiles each individual video flow that passes through the mobile packet core gateway and uses its machine learning algorithms to determine whether that flow is traversing a congested channel. It then flow-controls video to varying levels and time, depending on the degree of detected congestion, and efficiently aligns delivery of the video traffic to less congested moments while still providing adequate bandwidth to videos to maintain their quality. The result is less network latency and higher user throughputs while maintaining high-definition (HD) video. During this process, CUTO does not drop packets or modify data payloads in any way (see Figure 5).

FIGURE 5

How CUTO Increases Cell Capacity



Source: Cisco, 2020

The result is an optimized RAN, where mobile network operators can deploy fewer cells, on an ongoing basis, and absorb more traffic growth while meeting network quality targets. To summarize, CUTO provides the following benefits in congested cell sites:

- It increases average user throughput.
- It increases congested cell site capacity.
- It reduces scheduler latency.
- It maintains user quality of experience even when more users and more traffic share a cell.

The key outcome from an MNO perspective is that CUTO results in an improvement in KPIs utilized by MNOs in network capacity planning, such as UE throughput, scheduler latency, and network density, and delays more expensive capital investments in network capacity.

What Makes CUTO Compelling Compared with Alternatives?

Openness/Multivendor Interoperability

One of the key reasons why IDC believes CUTO is a compelling solution is its openness and its multivendor interoperability. Although CUTO integrates natively with Cisco's mobile packet core solution, it is inherently vendor and technology agnostic. It provides immediate gains across all RAN vendors, including Ericsson, Huawei, and Nokia. It is also cellular radio technology agnostic. It can work across cellular technology generations (3G, 4G, and 5G).

Operational and Implementation Differentiators

CUTO offers immediate cellular network capacity relief without the high complexity, extended lead times, and execution risk associated with traditional RAN hardware- or software-based network densification solutions. Through its advanced optimization, the solution can help deliver a better return

from RAN and mobile network investments. Specifically, what makes CUTO special from technical and operational perspective are the following features:

- **AI-based software-driven traffic optimization:** CUTO is a leading-edge AI-based pure software solution and involves no deployment of new hardware or cabling.
- **Faster time to deploy compared with alternatives:** CUTO integrates with the MNO's packet core P-GW function and can be up and running in an existing packet core deployment within hours.
- **Minimal disruption to existing infrastructure:** The CUTO software is lightweight and has minimal impact on deployed packet core CPUs.
- **Compatibility with TCP optimization:** Some providers also incorporate TCP optimization technology for RAN performance improvements. CUTO is compatible with TCP optimization.
- **Applicability to all traffic:** CUTO can work with both encrypted and unencrypted traffic, and it supports all traffic including UDP, HTTP, and QUIC.
- **Deployment flexibility:** CUTO can be deployed/integrated with Cisco's packet core StarOS P-GW, or it can be deployed standalone.

Economic Differentiators

CUTO is more affordable from both capex and opex perspectives compared with alternatives.

From a capex perspective, considering the solution can alleviate cellular network congestion without the deployment of any new spectrum or new hardware or the provisioning of new facilities, the advantages of the solution compared with its alternatives can be quite significant. These factors in turn also play a big role in lowering operating expense compared with the alternatives and the base case without the deployment of any cell site densification technology.

The customer case studies discussed in the section that follows highlight CUTO's benefits and differentiators as a solution.

Case Studies

Case Study 1: Spark New Zealand

Rugby is one of the most popular sports in New Zealand. At the time the Rugby World Cup was coming up in 2019, New Zealand-based mobile network operator Spark foresaw a need to improve its mobile network capacity with a view toward providing a superior fan experience for its mobile subscribers during the event while not compromising its existing services such as voice.

Spark recognized that to reduce traffic congestion in a radio access network, mobile operators have historically invested in additional radios and spectrum to increase mobile network capacity. This kind of buildout can cost millions of dollars and take nearly a year in time.

Spark decided to choose Cisco Ultra Traffic Optimization software solution to increase its radio site capacity and improve the quality of its mobile subscriber fan experience. In trialing this solution, Spark saw immediate gains in radio network capacity and performance because of stabilization of signal and maximization of per-subscriber throughput.

Specifically, during the trial, using CUTO, Spark:

- Increased user throughput by up to 30% across the day (up to 51% in peak hours)

- Reduced the RAN KPI samples <3Mbps by 33%
- Increased cell throughput on average by 2.6Mbps

Spark deployed CUTO nationwide across New Zealand ahead of the 2019 Rugby World Cup when the streaming of live games created peaks of network demand. According to Spark, CUTO was a key enabler of user experience of its mobile subscribers during the event. Importantly, Spark was able to delight its mobile data customers without negatively impacting its voice customers.

Case Study 2: Saudi Telecom Company

Saudi Telecom Company (STC) faced the problem of severe network congestion as a result of an unanticipated growth of video traffic. To alleviate the severe network capacity constraints, it was considering licensing of new spectrum bands with a view toward improving the quality of experience for its customers.

However, before the company could go ahead with licensing new spectrum and building network capacity around this new spectrum, it decided to trial CUTO. There were several reasons why STC chose CUTO for a trial. STC appreciated the fact that the software-only solution was cellular radio technology agnostic, it had minimal impact on packet core CPU performance, it was compatible with TCP optimization, it could work on all traffic, and it had deployment flexibility. While the trial was performed in conjunction with the Cisco Packet core, the fact that CUTO was interoperable with the packet cores of several other vendors was a positive.

The CUTO trial has shown encouraging results according to STC. STC has seen a per-subscriber throughput increase consistent with an overall cell-site capacity increase. STC is also pleased that CUTO is a significantly lower capital-intensive solution compared with the alternative of investing in a new buildout around new spectrum.

CONCLUSIONS

Consumer mobile video continues to strain cellular network capacity across the globe for mobile network operators. While operators continue to be capital constrained, they cannot afford to not invest in increasing their network capacity in order to maintain a superior customer experience in the face of continuing fierce competition in the industry. Traditionally, MNOs have either invested in new spectrum or new radio to maximize use of existing spectrum or used load balancing solutions such as SON. However, all these solutions are capital intensive, disruptive to existing operations, or take a long time to deploy.

Cisco Ultra Traffic Optimization is a software-based RAN optimization solution that offers an affordable, nondisruptive, and open solution to the cellular network densification problem and can help delay the investment in more capital-intensive solutions for the MNO. The experience of two operators – Spark and STC – attests to CUTO's efficacy in significantly improving cellular network capacity and hence user experience without breaking the bank.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com