Congestion Relief for Your RAN

How to manage the bandwidth explosion with Cisco Ultra Traffic Optimization

Everyone wants fast connections

Service providers face challenges on multiple fronts. Their networks must be able to handle ever increasing bandwidth demands and costs related to the tremendous increases in 4G/LTE usage along with the move to 5G. They must contend with faster speeds, more mobile connections, more mobile services, and more video. Keeping users connected and happy is particularly difficult because networking has its own version of the Pareto Principle with a slightly different ratio. The top 20 percent of mobile users generate 62 percent of mobile data traffic, and large continuous flows of data often are the result of streaming video traffic. In fact, according to the Cisco Visual Networking Index, mobile video is projected to account for 79 percent of total mobile data traffic by 2022.[1]

Cisco Ultra Traffic Optimization offers an attractive alternative to deploying costly spectrum and additional cellular radio gear because it optimizes large traffic flows on congested cells, maximizing the radio network performance where it is most needed.

Contents

Coping with massive growth

To compete and keep their businesses profitable for shareholders, service providers must invest in revenue-driven solutions and/or decrease capital expenditures (CapEx) and operating expenses (OpEx). However, the increase in users and bandwidth required in the next few years is daunting. According to the Cisco Visual Networking Index, there will be:

- More users: By 2022, there will be about 4.8 billion Internet users or 60 percent of the world's projected population (8 billion).
- Faster speeds: Globally, average mobile connection speed will increase three-fold from 2017–2022, from 8.7 Mbps to 28.5 Mbps.
- More mobile connections: The average global IP traffic by 2022 will reach 543 petabytes per hour. Wi-Fi and mobile-connected devices will generate 79 percent of Internet traffic by 2022.
- More video: The world will reach 3 trillion Internet video minutes per month by 2022, which is 5 million years of video per month, or about 1.1 million video minutes every second.[2]

As the statistics indicate, mobile data traffic growth is largely driven by increasing video consumption, which is encouraged through service provider unlimited data plans. Traditionally, many operators use throttling or data caps to control heavy usage. Not surprisingly, users accustomed to unlimited bandwidth don't appreciate these measures. To support growing demands, operators may need to add spectrum and deploy more radio cells, which can incur enormous CapEx and OpEx costs.

How the video explosion affects network densification

On a network, large data bursts and flows like video can take up a major share of total bandwidth over a period of time. The continuing increases in streaming video traffic affect everybody on the network. A single two-minute video consumes thousands of times more cell site capacity than the same amount of time browsing websites like Twitter or answering email.

Most consumer interactions (96 percent) represent approximately 30 percent of traffic by volume and are standard flows like web and app browsing, social media usage, music, emails, and maps. Only 4 percent of consumer interactions are large flows, which represent approximately 70 percent of traffic by volume (see Figure 1). Large flows are mostly streaming videos but may include be large file downloads and application and operating system updates.

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 such as 3 Mbps per user. One way to measure the density of a mobile network is using a formula: bps/user/km2. In other words, what bit rate can the cell site provide to how many users in a given geography? 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.

Often, 3 Mbps per user is used as a threshold for a service provider to upgrade its RAN capacity. If the cell isn't able to achieve at least 3 Mbps with 70-percent utilization and at least three active subscribers per transmission time interval (TTI), RAN vendors recommend adding capacity to the RAN. For many operators, this threshold is the one that triggers new capital expenditures.

In many networks, cell congestion is created from relatively few data sessions on a minority of cell sites. Typically, users are concentrated in the network and a minority of cells carry most of the traffic. The capacity is determined by how many users can share the channel and still get connection speeds to support their applications with a reasonable quality of experience.

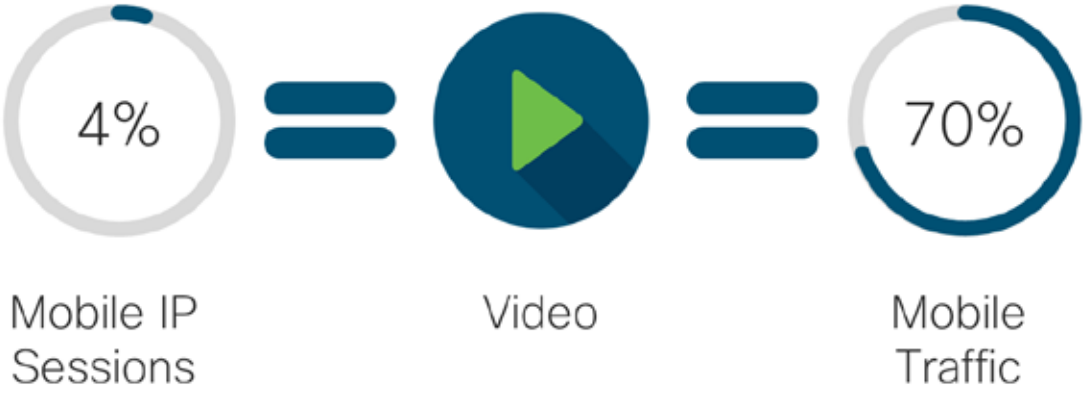
Congested cells generally have many concurrent large flows that are usually streaming video users. This small subset of users causes heavy contention for spectrum resources at a node and starves sessions of bandwidth. This situation reduces average connection speed for most users and drives the demand for new capacity to maintain an acceptable experience for users. These congested cells are the sites that will require densification and an enormous CapEx infusion as traffic grows year over year.

Traditional approaches to managing bandwidth and large flows

Streaming video is a growing phenomenon with new entrants that include Disney, Apple, and Amazon. However, adding more cells is expensive if you determine your RAN capacity requirements based on short temporary demand peaks in the network. In this situation, adding more capacity leads to a substantial underutilization of total RAN capacity. However, if you don't add capacity the user experience is degraded.

To manage demands, service providers often implement RAN load balancing using self-organizing network (SON) applications for dynamic load balancing (DLB) and intercarrier load balancing (ICLB) to replace manual efforts. Some providers also incorporate TCP optimization technology, which adjusts the network congestion avoidance parameters of TCP to increase performance.

Figure 1. The effect of video traffic



Video optimization is another approach to managing bandwidth challenges. However, the problem with current video optimization is that it uses techniques such as transrating and transcoding, which are obsolete because of encryption such as SPDY. It also requires vast amounts of hardware and processing power.

Throttling is another approach, but it degrades the overall network without offering congestion relief because the congested cells run below the flat throttle levels. Flat throttling reduces video quality for everyone because it affects non-congested cells in addition to the congested ones. Although throttling can reduce the overall traffic volume, the reduction doesn't occur in the congested cells, only in the open cells where video could have run at higher resolutions. With throttling in these open cells, the video quality is unnecessarily reduced to a lower quality. Throttling also increases the time a video spends on a channel, which drains battery and increases collisions with other traffic. In other words, throttling actually adds to network congestion while decreasing consumer satisfaction.

Getting smarter about congestion

Managing large flows and cell congestion is a digital communications issue that is most effectively resolved using software intelligence. By incorporating machine learning, software can optimize large flows, so it self-organizes into periods of time with surplus RAN capacity. This approach mitigates contention with non-large flows, which leads to increased speed and cell site capacity.

Cisco Ultra Traffic Optimization has a machine learning algorithm that optimizes all traffic to use existing RAN capacity more efficiently. This software-enabled technology offers the same results as adding more cells or spectrum to the RAN instantly without the cost, time, and complexity of adding new cells. Although new cells will always be required because of increasing traffic demands, Cisco Ultra Traffic Optimization decreases the number you need by using the existing capacity more efficiently.

Cisco Ultra Traffic Optimization is a more affordable alternative to purchasing new RAN equipment and spectrum. It offers immediate capacity relief without the high complexity, extended lead times, and execution risk associated with traditional RAN hardware densification. Through its advanced optimization, the solution can help deliver a better return from RAN and mobile network investments.

Cisco Ultra Traffic Optimization is vendor and technology agnostic. It provides immediate gain across all RAN vendors, including Ericsson, Huawei, Nokia, and across technology generations (3G, 4G, and 5G). Cisco Ultra Traffic Optimization also doesn't have the complexity, lead times, or execution risks associated with traditional hardware densification. Adding new spectrum or deploying new cells in congested areas like a dense urban market can easily take 18–24 months. By comparison, it takes only around two hours to implement Cisco Ultra Traffic Optimization. The solution is not only less costly, it also takes far less time. The solution can be enabled in Cisco Packet Core with less than a 3-percent CPU impact or operate as a standalone solution.

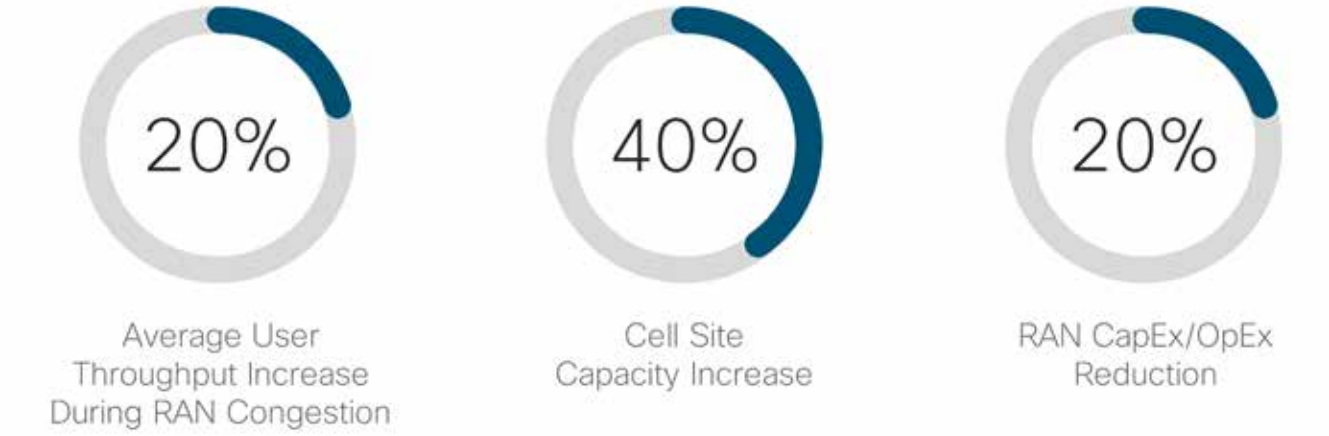
Savings that add up

Cisco Ultra Traffic Optimization can immediately provide RAN savings in both CapEx and OpEx and a capital efficiency multiplier year-over-year across the entire network of current and new deployments. The solution gives you the same network performance and capacity benefits as adding spectrum and cell site hardware to your networks.

When Cisco Ultra Traffic Optimization increases average user throughput, fewer cells fall under the high congestion levels that are defined by the RAN planning team. Cisco Ultra Traffic Optimization can increase the average user throughput during RAN congestion by 20 percent. The increase in user throughput provides an increase in cell site capacity of 40 percent, which leads to an estimated overall reduction in new RAN CapEx and OpEx of 20 percent (see Figure 2).

Cisco Ultra Traffic Optimization improves the quality of experience for all of the subscribers on the network. It increases data flows by increasing the average throughput per user as much as 30–40 percent in congested cells. It also improves the experience because the video experience isn't affected. There's no material impact to video start time, stall rate, or resolution. And unlike video optimization options, Cisco Ultra Traffic Optimization optimizes all traffic, whether it's encrypted or nonencrypted, video, or non-video. It increases subscriber connection speeds and cell capacity while providing a high quality (HD) video experience.

Figure 2. The reduction in new RAN CapEx and OpEx



How much can you save?

Here's a hypothetical example of the effect Cisco Ultra Traffic Optimization can have on CapEx and OpEx. Suppose that your network has 30 million subscribers and is operated with 200,000 carriers. If 2 percent of those cells are congested, you could save $1.2 Billion in CapEx ($300M) and OpEx ($900M) of new cells in five years by using Cisco Ultra Traffic Optimization.

How Cisco Ultra Traffic Optimization works

In each cell, software known as the "scheduler" is faced with the task of sorting through traffic and placing it into frequency and time slots for delivery over the air-link to the mobile devices. With video, the scheduler quickly becomes saturated and latency grows. The active user data sessions start backing off and slowing down. The performance of the cell site begins to crumble, and the capacity of the cell site is exhausted. Unmanaged video results in tremendous inefficiency and a reduction in capacity.

Video uses many concurrent resource blocks when it doesn't really need to, so the scheduler ends up allocating unnecessary bandwidth to video users at the expense of the other 90 percent of users. That 90 percent is adversely affected when the cell reaches high utilization and video users are in the mix.

The traffic load on cell sites varies from second to second, and Cisco Ultra Traffic Optimization reacts to real-time network congestion on a second-to-second basis. It proactively monitors a set of network metrics for cross traffic contention detection. By profiling congestion, it builds a statistical model of the data channel. That model is then used to detect congestion. The efficient usage of surplus network capacity manages the delivery of large flow while protecting other traffic from congestion. Managing these individual flows has the holistic effect of managing congested cells.

Cisco Ultra Traffic Optimization makes sure that network resources are available for all traffic flows when they need it. During times of congestion, large video flows are managed to provide cycles to starving smaller flows, which preserves the same or better video quality at congestion points.

With Cisco Ultra Traffic Optimization, these large video streams are adaptively flow-controlled at the packet core. The machine learning software begins to understand the environment and traffic conditions of the destination cell site and intelligently manages the flow of the video into the cell packet-by-packet. This removes the contention at the scheduler, frees up capacity, and increases the performance of the cell site and its capacity.

Cisco Ultra Traffic Optimization is compatible with TCP optimization and SON load-balancing technology. It doesn't monitor aggregate usage in a cell or cell congestion status. It considers each IP flow uniquely and analyses that flow to determine if a) it's a large flow and then b) if the large flow is experiencing congestion. It monitors all of the flows presented to it, and more closely monitors those flows that it identifies as large flows, looking for signs of congestion in only those flows.

Cisco Ultra Traffic Optimization is most useful in dense, congested areas. Some cell sites operate at very low physical resource block (PRB) utilization levels the majority of the time. In other words, they have little traffic and low numbers of active users. These cell sites tend to be for coverage and aren't located in dense urban and suburban areas. If a cell site has a very low volume of video users and therefore a low volume of traffic, Cisco Ultra Traffic Optimization provides little gain. However, these sites don't need and aren't the focus of densification in the network. Heavily used congested sites are where Cisco Ultra Traffic Optimization has the most value.

Achieving multidomain service orchestration

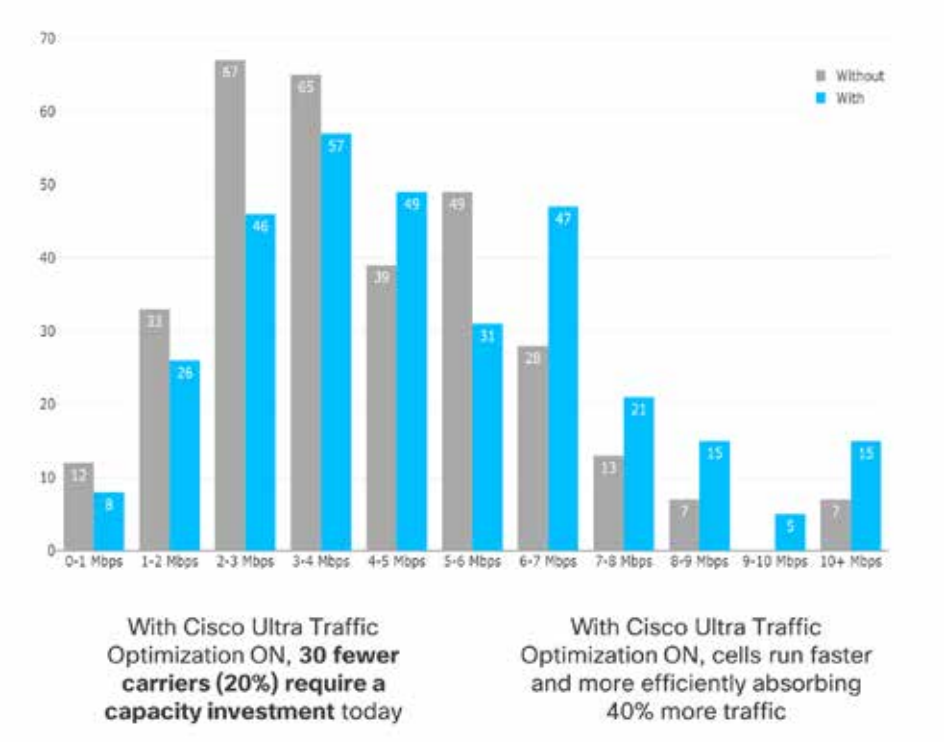
Cisco strives to deliver enhanced results from intelligent interworking of multiple solutions. Cisco Ultra Traffic Optimization working with Cisco Crosswork User-centric SON delivers end-to-end optimized and enhanced user experience. When paired together, you get multidomain service orchestration (MDSO). MDSO goes far beyond self-organizing network (SON) solutions and focuses on key performance indicators (KPI) at the radio. It uses quality of experience (QoE) and key quality indicators (KQIs) as additional input. This machine-learning solution pulls telemetry and analytics from several sources from the end device (IMSI) through the mobile core and policy. MDSO is set to monitor and optimize different service-level agreements (SLAs) for different services such as Internet of Things (IoT), video, and voice in an automated closed-loop manner. As a result, you extend the value and enhanced user experience from cloud-to-client. You end up with happier customers while reducing congestion and costs.

Cisco Ultra Traffic Optimization in action

Cisco Ultra Traffic Optimization has been validated in multiple networks through live trials that showed deploying the software had the equivalent effect of adding spectrum. Trials with similar results have been completed in North America, Latin America, and Asia. For example, in a European service provider, we found that Cisco Ultra Traffic Optimization was equivalent to adding spectrum. In this situation:

- 55 percent of users were on sites that averaged 8 Mbps of increased throughput with a peak of 15.61 Mbps.
- Ensuring an increase in site capacity across all these cells using hardware would require a spectrum overlay of 10 MHz.
- Without Cisco Ultra Traffic Optimization, the cell site capacity is exhausted quickly as video services continue to grow.

Figure 3. Example results from using Cisco Ultra Traffic Optimization



Learn more

Unlike other solutions, Cisco Ultra Traffic Optimization can be used for all types of traffic and flows and all types of RAN technology. It improves user experience as it increases subscriber connection speed and cell capacity using software while providing high quality (HD) video. It also reduces RAN-related CapEx and OpEx. Cisco has tools to help you calculate the potential return on investment in your situation.

To find out how Cisco Ultra Traffic Optimization can help you beat the bandwidth explosion and manage large flows, visit www.cisco.com/go/mobile

In another trial, a 10-MHz eNodeB that was carrying 15 Mbps can now carry 22 Mbps, which is the equivalent to adding a 4-MHz spectrum band to every cell in the network. Network-wide software deployment also now typically is completed in one hour.

In Latin America, trials showed a significant improvement in RAN KPI throughput. Cisco Ultra Traffic Optimization was measured to increase bits/second per user by 20 percent in the trial. The traffic carrying capability of a cell site increased, which in turn meant that the download data volume threshold could be increased for capacity planning purposes.

Cisco Ultra Traffic Optimization demonstrated:

- Average all-day improvement of 20 percent.
- Peak improvement of 55 percent in user throughput RAN KPI.
- Both commensurate with a 40-percent cell capacity gain.

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html
2. Ibid.