WHITEPAPER

Designing a scalable network to unleash the true potential of 5G

Published by

Executive summary

5G is not just another 'G' in the evolution of cellular radio networks – it is so much more. 5G promises incredible advancements over 4G, including data rates of up to 10 Gb/s, latencies as low as 1ms, densities of up to 10 Tb/s and the ability to connect up to 1 million devices per square kilometer.

However, the secret to delivering that lies not only in the radio antennas and base stations, but the core transport network that sits behind it. 5G base stations can't deliver superfast speeds and ultra-low latencies if the core network is gummed up with congested data traffic. The core network must be congestion-free, else the whole value proposition of 5G falls apart.

Philippines incumbent operator PLDT has big plans for 5G, which will enable its cellular arm Smart Communications to build innovative services and grow revenues. For 5G to fulfil that potential, the transport network must be free of congestion, but standard mobile architectures cannot guarantee zero congestion in all use-case scenarios. PLDT needs a new network architecture designed to eliminate bottlenecks once and for all and provide a superior customer experience at virtually infinite scale to support the expected proliferation of end points and new services.

Luckily, the technology to do this exists: it's all a matter of using the right components to build the right architecture. This paper looks at how PLDT and Cisco are designing a new and scalable network architecture that can unleash the true potential of 5G.

Introduction

PLDT has had its sights firmly set on 5G even before the technology was standardized. But while 5G's benefits of exponentially faster data speeds, higher capacity and low latency are certainly attractive, PLDT understood that those benefits could only be realized if the core transport network was capable of handling the inevitable explosion of data traffic that would follow. Any congestion in the transport network would impact the customer's 5G experience no matter how fast the wireless link between the device and the base station.

Ironically, PLDT first became aware of this need for a congestion-free network several years ago when it rolled out 4G. Once LTE traffic began to take off, it became clear quickly that PLDT's transport network wasn't designed to handle even 4G traffic demands, let alone 5G.

Operators have traditionally planned voice and data network capacity based on oversubscription models that assume that, statistically, not everyone is going to be using the network all the time or use up all the available capacity during peak hours. However, these models don't work well in the era of "always on" data connectivity.

Meanwhile, PLDT's customer surveys revealed that customers value service reliability over speed. This means it's paramount to be able to deliver not just a good quality experience, but a consistent one – customers see no value in a 1 Gb/s broadband connection if the speed drops to 200 Mb/s as soon as one of their neighbours starts streaming 4K videos.

This will be especially true with 5G. A transport network designed to be congestion-free end to end – and with 5G's particular characteristics in mind – is absolutely essential to ensuring 5G can deliver that consistent, reliable experience that customers expect.

This required a paradigm shift in network design from average-load planning to peak-load planning. Thus, PLDT decided to redesign and rebuild its transport network architecture not only to meet exploding demand for 4G traffic, but prepare for the enormous capacity needs of 5G. PLDT also had to factor in architectural requirements that are unique to 5G, such as massive scalability, indoor coverage and mobile edge computing (MEC).

Ingredients of a congestion-free network

A congestion-free transport network requires several key new technologies and innovations that are just now becoming available.

High capacity routers

The first element is faster and more scalable routers that not only offer capacity more than four times that of the latest and largest platforms, but also do so using a fraction of the footprint and power. This latter characteristic is crucial, because otherwise the result is a faster router that burns more power and takes up more space in the data center.

There are several technologies that enable this:

- **400G coherent optical technology**, which is being rapidly commoditized due to demand from OTT, web scale and service providers
- **Zero Density trade-off pluggables with digital coherent optics (DCO)** – an OIF standardized form factor in which 7nm processes enable Cisco to not only fit more 400G optical interfaces into the same footprint, but to actually make the footprint smaller.
- **High-performance ASICs** – a router of congestion-free caliber requires a highly scalable platform providing sufficient backplane capacity. To this end, Cisco has developed a high-performance ASIC with a large FIB (Forwarding Information Base) size and deep buffers to enable a platform in excess of 10Tbps capacity in one RU (Rack Unit), providing a zero-blocking router architecture.

Converged SDN architecture

Another key element is a **converged SDN transport architecture** to simplify and optimize the 5G network.

Simplification is key because typically, today's networks are built in isolation of one another – there are separate networks dedicated for different purposes, such as enterprise services or mobile backhaul. These siloed networks must be converged into an **SDN transport architecture** that ties everything together onto a single pane of glass to make sure that the network is optimized around services instead of infrastructure.

SDN capabilities with orchestration and automation are needed to simplify operational models and bring new operational efficiencies reducing time to market of new services and bandwidth scale-out requirements. This also contributes towards providing a superior customer experience, while increasing network utilization and reducing TCO.

This in turn means that routing protocols also need to be simplified, which brings us to another key element of the congestion-free network: **segment routing (SR)**.

SR is an evolution of MPLS VPN that simplifies the network control protocol stack, enabling a unified, end-to-end, policy-aware network architecture from the data center servers to the IP transport network of service provider networks. This increases scalability and enables enforcement of strict performance guarantees with application-level granularity.

SR also forms the foundation of end-to-end network slicing, a crucial element of 5G's value proposition, particularly for enterprise customers. SR also enables cost-efficient network scaling via its superior traffic engineering capabilities that ensure efficient use of network resources.

IP+Optical

Current mobile network architectures have both IP and optical networks – where the optical layer connects all the IP routers. But these networks are built and managed completely separately.

Complicating things is the fact that each layer has different technology refresh cycles, which results in uneven capacity upgrades. For example, if a service provider with a 100G optical transport network needs to upgrade its routers to 400G, the optical network needs to be upgraded to support the router upgrade.

To ensure both the IP and the optical layer are upgraded at the same pace, Cisco's solution is to converge these layers into a single "IP+Optical" layer. This creates a more scalable architecture and reduces functional overlap.

Integrating optics onto the router eliminates the need for a separate optical layer and enables service providers to maintain the same pace of evolution for both IP and optical. A converged IP and optical layer also enables service providers to add capacity where they need it rather than forklift the entire network.

Designing the congestion-free architecture

With the elements of the congestion-free 5G transport network identified, the next step is knitting it all together into a viable network architecture. However, this requires a break from conventional thinking.

Traditionally, service providers connect routers using a ROADM-based DWDM architecture. There are usually two approaches to this:

1. Hollow core architecture – a full-mesh ROADM network that allows any router to be connected to any other router on the optical network. This architecture has more optical DWDM ports and fewer router ports (making it possible to save on router capacity and interfaces). Downsides include limited flexibility, longer distances and high opex from maintaining both an IP and optical network, while capacity and scalability are restricted by Shannon's limit. For 5G specifically, a hollow core means traffic is tromboned through the core, which impacts latency, particularly for east-west traffic – that's virtually a dealbreaker in the 5G world.

2. Optical bypass architecture – this architecture attempts to leverage the advantages of IP and optical by routing east-west traffic directly between router nodes, but with only the traffic heading towards a destination dropped at its respective router. All other traffic goes through the core nodes, which saves ports on the IP router. The limitations of this architecture are similar than the hollow core architecture. Also, optical bypass made economic sense in an era where IP ports were much more expensive than optical ports, but the commoditization of 400G optics mentioned above is also driving down the cost of IP ports, which eliminates a primary rationale for this approach.

An alternative approach that is more suitable for the congestion-free network is a **hop-to hop-architecture** that gets rid of the ROADMs completely.

Hop-to-hop architecture, explained

In a hop-to-hop architecture, all optical traffic is dropped into a router at every hop, reducing fiber spans to the smallest possible distance, thus allowing for the maximum capacity possible as defined by Shannon's limit. By maximizing fiber capacity, more lambdas can be added by filling more ports on the routers, or even adding more routers if needed.

This architecture allows the service provider to deliver maximum capacity to each cell site, which in turn maximizes scalability and the customer experience. It also futureproofs the architecture by eliminating the need for forklift upgrades. Meanwhile, segment routing simplifies the underlay, making it easier for automation platforms to support advanced use cases like traffic engineering and bandwidth on demand.

The combined IP+Optical layer also gives the optical transport network the flexibility of an all-IP network. 5G needs that level of flexibility in terms of where the service provider can place services and data centers at the edge to support low-latency use cases. Segment routing can largely automate the process of deploying micro data centers to serve, for example, manufacturing plants deploying private 5G across their campuses.

Perhaps the most attractive benefit of a hop-to-hop architecture: it's much more cost-effective. Converging the IP and optical layers saves a considerable amount of capex and opex. While the amount of savings depend on which traditional ROADM-based architecture you compare it with (and will of course vary from one operator to another), based on modelling of real Service Provider networks for the different scenarios, Cisco sees potential relative capex savings in the order of 40% over a Hollow Core network and an approximate saving of 34% over an optimal bypass network.

For PLDT, a key attraction of a hop-to-hop architecture is that it best serves 5G's latency requirements. Latency is essentially a product of physical distance – a general rule of thumb is 1ms of latency equals roughly 100 km. Thus, to achieve 1ms latency, the data being transmitted must be hosted within 100 km of the customer.

This is problematic with current cloud-based architectures because the data could be hosted one block away or halfway around the world. The reality of the Philippines is that most content consumed online is hosted overseas, typically in the US. The latency between the US to the Philippines is 180ms, which is clearly unsuitable for 5G, thus necessitating the need for MEC.

Even when content is hosted within the country, the Philippines is a lengthy archipelago – the distance between north Luzon and south Mindanao is substantial enough to create considerable latency. The transport network must be designed to avoid tromboning data traffic. A hop-to-hop architecture ensures that traffic routes are as direct as possible, with the shortest distance possible and with as few hops as possible.

Challenges to consider

Like many incumbents, PLDT had to decide what would become of its legacy gear in implementing this new transport architecture. The operator ultimately decided to build a new network in parallel to the existing one, then deploy all-new services on the new network, whilst migrating existing services to the new network at its leisure.

This isn't as easy as it sounds. For one thing, it's painstaking work – migrating traffic from the legacy network to the new network means changing the routes for each network element, of which there are hundreds of thousands. Also, it's the equivalent of open heart surgery because all services must continue to run seamlessly as the migration is executed.

PLDT faced another notable challenge: the condition of its physical fiber. Between the earthquakes and typhoons that frequently batter the archipelago, fiber breaks are a regular occurrence in the Philippines. The resulting splices increase attenuation of PLDT's long-haul links, particularly when migrating from 10G and 100G to 200G and eventually 400G. PLDT had to refurbish its fiber to support the new transport architecture.

An additional challenge of building a congestion-free network that's unrelated to technology is the organizational mindset. Put simply, converging the IP and optical networks also means converging the teams that run them, which means people with different technical skills must learn to work together.

However, as luck would have it, this was not a specific issue for PLDT, which several years ago converged its combined its mobile transport network teams – covering wireless, IP and optical – into one organization.

Conclusion

The arrival of 5G has ushered the mobile operator industry into an era of not only unprecedented revenue opportunities, but also unprecedented transformation from the inside out, and from end to end. 5G will trigger a proliferation of new end points and services generating exponential growth of user traffic. This prospect has in turn triggered a rethink of how transport networks are designed, where the key metrics are no longer speed and latency but a consistent and reliable customer experience.

Having realized this early, PLDT has architected a new transport network capable of preventing or alleviating congestion in the core by leveraging new technological innovations in ASICs, digital coherent optics, high-capacity router platforms and SDN to converge the IP and optical layers.

Combining these elements enables a simplified hop-to-hop architecture that not only eliminates the bottlenecks that cause congestion, but provides the massive scalability that 5G requires and ensures the consistently reliable experience that customers expect. Result: a truly congestion-free transport network – and a cost-effective one at that, with reduced cost-per-Gigabyte and a lower TCO.

Cisco (NASDAQ: CSCO) is the worldwide technology leader that has been making the Internet work since 1984. Our people, products, and partners help society securely connect and seize tomorrow's digital opportunity today.

Discover more at www.cisco.com/go/5g-transport and follow us on Twitter at @Cisco.

Produced by the mobile industry for the mobile industry, Mobile World Live is the leading multimedia resource that keeps mobile professionals on top of the news and issues shaping the market. It offers daily breaking news from around the globe. Exclusive video interviews with business leaders and event reports provide comprehensive insight into the latest developments and key issues. All enhanced by incisive analysis from our team of expert commentators. Our responsive website design ensures the best reading experience on any device so readers can keep up-to-date wherever they are.

We also publish five regular eNewsletters to keep the mobile industry up-to-speed: The Mobile World Live Daily, plus weekly newsletters on Mobile Apps, Asia, Mobile Devices and Mobile Money.

What's more, Mobile World Live produces webinars, the Show Daily publications for all GSMA events and Mobile World Live TV – the award-winning broadcast service of Mobile World Congress and exclusive home to all GSMA event keynote presentations.

Find out more www.mobileworldlive.com