Monetize Immersive Streaming Services with a Virtual CDN

Deploy scalable media caching on VMware Telco Cloud Infrastructure

AT A GLANCE

- Elastic and scalable edge caching solution for delivering high-performance, low-latency quality of experience (QoE) services with elastic scaling of virtualized edge resources and dynamic composition of CDN services

- Cost-effective for reducing backhaul traffic and managing traffic spikes with virtualized infrastructure, dynamic virtual resources sharing, and idle resources utilization

- Centralized management for deploying and managing distributed edge architecture with a consistent and horizontal hybrid cloud platform across any network and any cloud

- Higher operational efficiency and service velocity for deploying and operating in thousands of remote CDN nodes with a multilayer cloud automation and operation platform

- High-performance, high-throughput capacity for supporting a large number of simultaneous 4K and ultra-HD video streams with a VMware NSX® managed virtual distributed switch (VDS) in EDP mode, hardware acceleration, and TLS termination

- Superior quality and video delivery with low latency, low jitter, deep edge caching, and smooth video playback under unpredictable network conditions

- Monetization opportunity as edge infrastructure as a service (IaaS) or platform as a service (PaaS) by leveraging common edge infrastructure

Introduction

Communications service providers (CSPs) weighing how to deal with surging volumes of Internet traffic will find the best, most cost-effective solution also opens new opportunities to drive monetization across B2C and B2B markets.

This is the strategic outlook for CSPs who modify their consolidated edge infrastructures to support a virtual content delivery network (vCDN) that can execute the capabilities of the joint solution offered by Intel, Varnish Software and VMware. By virtualizing CDN deeper into their networks, CSPs can dynamically allocate edge compute, storage, and bandwidth resources to a CDN and other functions as needed in the space-limited confines of deeply placed micro data centers.

First and foremost, the ability to automatically and dynamically allocate edge resources when and where is required gives operators an opportunity to future-proof networks against an ever swelling and constantly fluctuating tide of video and other Internet traffic. This is mission critical.

The opportunity

Global IP traffic volumes registered in 2017 are projected to nearly triple by year-end 2022, jumping from 122,000 petabytes to 396,000 petabytes monthly with the video share of the flow going from 75 percent to 82 percent[1]. The breadth of the swings in traffic fluctuations is growing as well, with peak rates increasing fivefold over that period compared to a fourfold increase in the average traffic rate.

These swings have been exacerbated by the growing share of traffic consumed by sports, news and other live content that can drive traffic to unpredictable heights at a moment's notice. On-demand traffic, which accounted for 81 percent of all IP video traffic in 2018, continues to increase, but live traffic is growing even faster on a pace projected to reach 238 exabytes by 2024, equating to 52.5 percent of the projected 453EB total video traffic that year[2].

By taking advantage of vCDN architecture to address these needs, CSPs will be able to accommodate the flow patterns of current and emerging consumer and commercial video, the Internet of Things (IoT), extended reality (XR), and other traffic categories. They'll also be able to make processing power available for the many other applications that edge facilities may be called on to perform.

This is a growing list that includes dynamic ad insertion, per-stream enhancements to user experience, execution of forensic watermarking, and much else. And, with the vCDN in operation, CSPs delivering 5G services will be able to execute data processing functions in close proximity to endpoints as required by a new generation of ultra-low-latency applications.

Moreover, a highly elastic distributed intelligence architecture creates new B2B service opportunities. CSPs can carve out dedicated uses of their network resources

VIDEO ON DEMAND (VoD)
- Provide on-demand video content to users over HTTP(S) and HTTP/3

- Scale easily to optimize resource utilization

- Optimize the delivery NUMA balance

- Can serve 50,000 connections per 100Gb connection, and multiple simultaneous streams from a large library

- Provide TLS/SSL termination along with a web application firewall (WAF) and distributed denial-of-service (DDoS) prevention

STATIC CONTENT WEB AND APPLICATION DELIVERY
- Provide web objects and file downloads over HTTP(S) and HTTP/2

- Enable reverse proxy web caching

- Maximize cache hit rate to reduce backhaul load and deliver quickly

- Optimize large numbers of secure connections of small file/objects

- Can serve 100,000 connections per 100Gb connection

- Provide TLS/SSL termination along with a WAF and DDoS prevention

LIVE VIDEO STREAMING
- Provide live IP video streaming to different users over HTTP(S) and HTTP/2

- Provide high QoE independent of bandwidth experience by each user

- Can serve 50,000 connections per 100Gb connection

- Provide TLS/SSL termination

- Reduce jitter and interruption for livestreams

for wholesale CDN offerings, deliver dynamic advertising support to other providers of OTT video services, and stand up dedicated private networking domains for enterprise customers.

The challenge

While all these opportunities have strong appeal, CSPs' most immediate challenge is to alleviate the pressure that traffic flows are exerting on CDN processes housed within their own infrastructures. With traffic spiking at the rates noted in the previous section, this is a challenge that cannot be met with legacy CDN technology without incurring untenable, recurring expansion expenses.

Most CSPs have overlaid their CDN architectures across existing facilities. Origin servers and caches operate in central locations either on purpose-built hardware or, increasingly, on general purpose servers in private cloud data centers with occasional use of public cloud resources as well. Video encoded and packaged for live or on-demand distribution is streamed out either directly or by way of mid-level cache locations to edge cache clusters typically positioned in telco central offices (COs) or cable headends (HEs).

CSPs must be sure there's always enough processing and storage capacity to handle the peak levels of users' streaming requests from across the entire area served from the CO/HE. In cases where purpose-built hardware is used, this requires ongoing investments in capacity that sits unused a large share of the time.

Further complicating matters is the fact that the CDN edge processing and storage requirements differ significantly between live and on-demand content, requiring different types of purpose-built components for each mode. Live content needs a small to moderate amount of storage capacity with high endurance and high usage volatility, while on-demand video requires much higher storage capacity requirements on a more consistent basis.

But even if the current edge facilities have been virtualized to enable a more fluid shared-use model for unused capacity, evolving service enhancements together with the increasing traffic flows make it hard to accommodate all the per-stream processing requirements for the entire local population of users. There's simply not enough space in these facilities to keep adding servers.

The CDN embodies a far more heterogeneous mix of workflow profiles than prevailed in the early years of video streaming. For example, the requirements for transactional e-commerce security with the transfer of stored content to end users as enabled by Transport Layer Security (TLS) has driven TLS processing to multi-100Gbps speeds. This requires higher CPU core counts and high levels of nonvolatile memory express (NVMe) storage capacity.

Moreover, CSPs must have the caching and processing capacity to handle ever more versions of content as 4K UHD/HDR formatting takes hold and networks begin delivering content formatted for virtual reality and other XR experiences. And there will be ever more per-stream storage and processing capacity required for video transcoding, packaging, and manifest manipulation associated with personalization of services and the implementation of edge-based dynamic ad insertion, which will also require digital ad insertion (DAI) functionality at the edge.

The solution

CSPs need to adopt a more cost-effective and agile approach to addressing these challenges and opportunities. They can do this by consolidating CDN service delivery on a scalable, multipurpose vCDN platform that orchestrates storage and distribution across distributed edge locations closer to end users. In doing so, operators can reduce local streaming workloads to manageable proportions with the versatility to

BENEFITS OF VIRTUALIZING EDGE CACHES ON VMWARE

- Virtualization increases agility and reduces the total cost of ownership (TCO)
- Disaggregation and abstracting hardware into a software layer increases agility, scalability and manageability
- Applications and servers can be upgraded at any time without any impact on service availability

ADVANTAGES OF RUNNING VARNISH EDGE CACHES ON VMWARE

- Multi-cloud flexibility increases agility for different deployment scenarios and business models
- Centralized management increases operational efficiency for deploying and managing distributed edge services with consistent infrastructure, automation and operations from the core to the cloud and the edge
- Multilayer automation accelerates time to market and increases the scalability, repeatability and operation efficiency at the CDN edge sites with VMware Telco Cloud Automation™
- Macro-segmentation increases revenue opportunities and protects multitenants in co-location or revenue-sharing business models
- An NSX managed VDS in EDP mode provides superior network performance and elastic scaling across a VMware vSphere® cluster
- End-to-end, real-time service assurance simplifies monitoring and performance management with VMware Telco Cloud Operations™
- Intrinsic security is built-in, unified and context-centric to accelerate risk identification, prevention, detection and response to threats with the right context and insights

use capacity as needed to handle fluctuating levels of traffic and to dynamically allocate resources to other applications.

Given these benefits, it's not surprising that the vCDN imperative has gained acceptance among CSPs worldwide. A recent global survey of operators found that 56 percent expect to be operating vCDNs by year-end 2021 with another 20 percent saying they'll get there by year-end 2022[4].

As the vCDN concept gains currency among CSPs, they need to ensure that the virtualization platform they choose can meet all requirements essential to a future-ready networking environment. Those capabilities are reflected in a joint solution that combines the versatility, scalability and reliability of the VMware Telco Cloud Infrastructure™ instantiated on Intel Xeon® Scalable processors with the comprehensive range of vCDN applications enabled by the CDN toolkit supplied by Varnish Software.
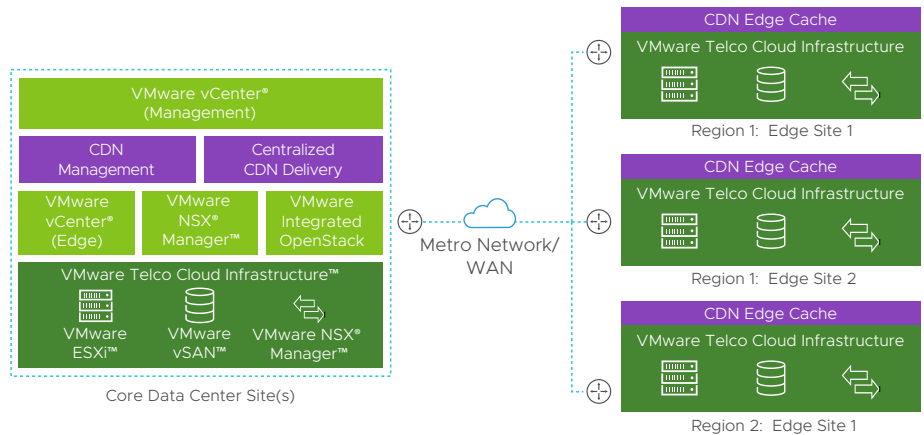

FIGURE 1: Centralized management and dynamic edge resources and services.

VMware Telco Cloud Infrastructure is a consistent and horizontal hybrid cloud platform that runs any application across any network, from the core to the edge of the network, and on any cloud. It supports centralized management and distributed edge architecture. It enables CSPs to leverage and optimize their infrastructure, and delivers consistent deployment for virtualized or container-based infrastructure supporting virtual network functions (VNFs) and containerized network functions (CNFs), regardless of where you deploy them, and be able to reconfigure and reallocate their infrastructure resources as needed. It reduces siloed operations, lowers TCO, and optimizes the network resources to scale.

By leveraging a VMware NSX managed VDS in Enhanced Data Path (EDP) mode with elastic resource scaling across vSphere clusters, VMware Telco Cloud Infrastructure delivers superior data plane performance, continuous service availability, and simplified management to the edge sites.

For media caching in a distributed and disaggregated edge architecture, data plane workloads can be deployed in data plane-intensive clusters. Use the NSX managed VDS in Enhanced mode feature to assign dedicated CPU resources to improve network performance for meeting low latency and higher throughput requirements at the CDN edge caching sites if, or when, it is needed.

In addition, CSPs can leverage VMware Telco Cloud Automation to improve operational efficiency and service velocity by delivering new network functions and services on demand and at scale across multiple clouds. VMware Telco Cloud

VARNISH SOFTWARE[3] EDGE CACHE

It is a caching layer between origin servers and clients/user devices with web browsers/video players. Content caching decreases page load times and maintains high-performance HTTP/HTTPS content delivery for complex, personalized and dynamic content. It is easier to scale based on demand and more responsive for low-latency services.

DESIGNED FOR EXTREME PERFORMANCE

• Approximately 200Gbps from a single node

• Sub-millisecond time to first byte

• 100,000 concurrent users per node

• Native, high-speed TLS termination

MINIMIZED BACKHAUL WITH INTELLIGENT MEDIA CACHING

• Share cached content across nodes

• Combine similar content requests

• Predictively load content into cache

MOVE CONTENT CLOSER

• Multi-terabyte edge storage (more than 100TB VOD libraries)

• API-based content replication

• Dynamic cache placement

Automation provides network services orchestration, such as vCDN, network functions lifecycle management, and multi-cloud management. It automates the instantiation of CDN cache based on the demand and policy by automatically allocating resources in each location to applications based on resource availability and operator priorities at any moment in time. Scaling is further aided by automatic instantiation of new vCDN edge nodes as capacity needs dictate. VMware Telco Cloud Automation also supports zero-touch provisioning of the virtual infrastructure, inclusive of the container-as-a-service (CaaS) layer and automation of Kubernetes cluster management, and Day 0 and Day 2 operations. It provides the scalability, repeatability, velocity and operational efficiency needed to serve remote edge sites. With these capabilities, CSPs can dynamically deploy, redeploy and upgrade new services across hundreds, thousands and even hundreds of thousands of edge nodes. And they can add, remove and reallocate capacities and capabilities for specific service groups across the entirety or any subset of edge service areas in a reliable and automated fashion.

Whatever the scale of operations might be, VMware Telco Cloud Operations delivers carrier-grade performance with real-time service assurance by correlating the health of the virtual services with the underlying infrastructure and managing the network accordingly. VMware Telco Cloud Operations simplifies the monitoring and management of a complex CDN. By utilizing integrated performance analytics aided by machine learning and automated fault analysis, the platform provides comprehensive visibility and automation that triggers proactive responses to emerging issues before they become disruptive to operations.

By leveraging the widely deployed vSphere hypervisor and the NSX network security virtualization platform, VMware, Intel and Varnish Software offer a pre-integrated solution ready for large-scale vCDN deployments for media caching and streaming in telco cloud environments.

With hardware acceleration powered by second-generation Intel Xeon® Scalable processors that leverage the open source Data Plane Development Kit (DPDK), the joint vCDN solution delivers high performance with fast packet throughput, space and power savings, and other advances that have been optimized for CDN processing, storage and networking. Benefits include optimum core counts; instruction sets for cryptographic, AI and video processing; and CDN-optimized CPU SKUs, enabling high throughput and low latency for CDN workflows.

Intel storage advances include high-capacity, high-I/O solutions that enable heavy write access without sacrificing read for video-on-demand (VOD) applications, aided by intelligent caching profiles supported by user behavior analysis. For live video delivery, where content is cached to facilitate large client stream volumes in conformance with low-latency service-level agreements (SLAs), Intel Optane™ persistent memory supports high-density, high-endurance media for comparable low workload latency characteristics.

With Varnish caching and streaming software running on VMware Telco Cloud Infrastructure, the joint solution enables a highly scalable vCDN for low-latency, high-capacity content delivered to the edge, and meets any large-scale, dynamic media streaming needed to defeat the capacity gap. It opens up new opportunities for serving both consumers and enterprise customers the content they need on any device and any cloud. It enables cost-efficient, optimal utilization of edge resources, so CSPs have an elastic scaling response to traffic spikes and shared processing across CDN and non-CDN applications.

Varnish edge caching nodes offer high throughput at very low latency, meeting requirements for delivering high-quality live, OTT and on-demand video with excellent QoE. High-capacity caches can serve up to 100,000 requests per second, while edge storage of more than 100TB maximizes cache hit rates to reduce backhaul and keep

latency low. Varnish caching nodes are also capable of TLS termination at high speeds.

In lab testing with 10 Varnish-Plus instances, the joint solution delivered high performance with an aggregate throughput of more than 190Gbps video content from a single server in a virtualized environment, nearly the limit of two 100GbE links.

Value proposition and summary

Figure 2 shows the advantages of the vCDN solution supported by VMware, Intel and Varnish reduce TCO for CSPs' operations in the Internet services domain.

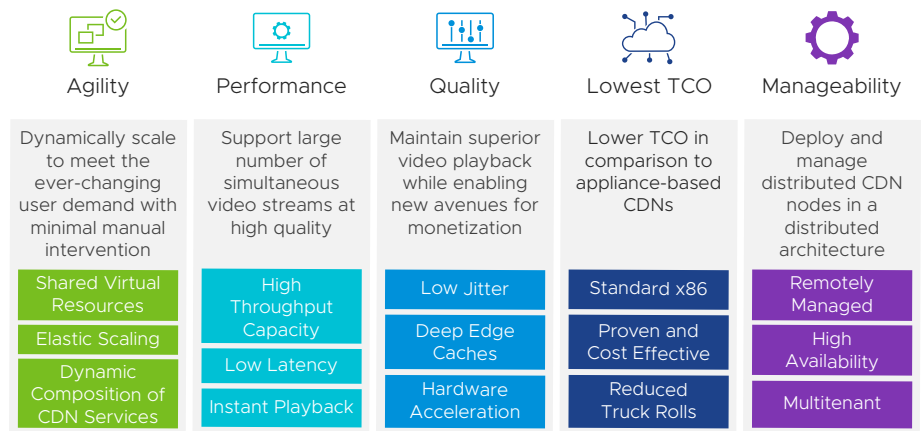| Agility	Performance	Quality	Lowest TCO	Manageability
Dynamically scale to meet the ever-changing user demand with minimal manual intervention	Support large number of simultaneous video streams at high quality	Maintain superior video playback while enabling new avenues for monetization	Lower TCO in comparison to appliance-based CDNs	Deploy and manage distributed CDN nodes in a distributed architecture
Shared Virtual Resources	High Throughput Capacity	Low Jitter	Standard x86	Remotely Managed
Elastic Scaling	Low Latency	Deep Edge Caches	Proven and Cost Effective	High Availability
Dynamic Composition of CDN Services	Instant Playback	Hardware Acceleration	Reduced Truck Rolls	Multitenant

FIGURE 2: Benefits of the joint VMware/Intel/Varnish vCDN media caching solution.

Costs incurred with instantiations of new edge facilities are more than mitigated by the savings that result from the use of virtualization technology to optimize dynamic, system-wide orchestration of resource allocation across core and edge facilities.

There simply is no more cost-effective way to accommodate traffic surges, mounting cache storage loads, and multiple application requirements with adherence to the rigors of carrier-grade performance.

The TCO advantage will only increase over time in comparison to what it would cost to equip the less dispersed purpose-built infrastructures of the past to accommodate soaring streaming traffic volumes and new demands on hardware resources emerging with new video formats, new service categories, and opportunities to improve user experience. Moreover, sizeable TCO reductions will result from the monetization that comes with dynamic per-stream advertising and B2B opportunities related to offering CDN services and support for edge-based applications to OTT service providers.

1. Cisco Systems. "More IP Traffic in the Next Five Years than in the History of the Internet." November 2018.
2. Broadband Technology Report. "Live video, low latency to dominate CDN traffic." July 2019.
3. Varnish Software. "Edge Cache and media streaming features are supported and tested by Varnish Software."
4. Light Reading. "How to Build a Cloud-Native CDN and Bring Scalability to the Edge." November 2020.

vmware®