Kollective

Control and Adapt Mechanisms Drive Successful Video Delivery Within Global Enterprise Networks

KOLLECTIVE SD ECDN, A FUTURE-PROOF VIDEO AND LARGE-FILE SOFTWARE DELIVERY SYSTEM

TABLE OF CONTENTS

Importance of Control and Adapt

Much like how the principals of natural selection drive the evolution of animal species, the Control and Adapt mechanisms of Kollective's software-defined enterprise content delivery network (SD ECDN) ensure video and other large-file content withstand the complex and dynamic environments of today's global enterprise networks. In the term "survival of the fittest", "fittest" is meant as the one best suited for the immediate environment, not always the strongest, fastest or smartest. In this relationship between the individual and the environment, the environment plays a crucial role, and the individual must find a way to adapt to it in order to survive and prosper.

Likewise, in the world of information technology with its vast array of diverse, global enterprise network environments, software solutions must be capable of adapting to, and exerting a measure of control over, the unique circumstances present in each and every one of these networks.

After 15 years of experience delivering video and other large-file content to over 3.5 million users on every continent, we have learned this to be true: no widely-dispersed enterprise network is exactly like any other and therefore enterprise content delivery technology that is unable to adapt within these different and changing environments will not survive to serve its purpose.

SD ECDN Defined

SD ECDN is a software-based overlay system that orchestrates enterprise network endpoints (PCs, Macs, VDI terminals and mobile devices) into an adaptive, continuously optimizing, fully-distributed content distribution grid. Its formation and operation are fully software-defined, providing the flexibility, agility, and central control commonly afforded by software-defined systems. Specifically, SD ECDN stands for Software Defined Enterprise Content Delivery Network. Software Defined is a term that is increasingly used to describe solutions that deliver network capability in software that may have historically been delivered via hardware devices. The ability for Software Defined solutions to be delivered in a fraction of the time, for a fraction of the cost of their hardware based predecessors has led to the rapid adoption of a wide range of Software Defined technologies in today's enterprise environment.

The Kollective SD ECDN consists of a set of cloud-hosted control and origin servers and a small software agent deployed on employee devices throughout the company. The central servers and agents collectively form an adaptive, distributed content delivery and caching system to ensure that upwards of 99% of content is delivered via controlled, localized, east-west traffic instead of across the more constrained wide-area and internet gateway links often referred to as north-south traffic.

SD ECDN technology leverages existing infrastructure, notably storage and network bandwidth on end-user devices. It is extremely secure and delivers content via a multi-layered crypto-protected mesh that dynamically adapts to network and other consequential changes.

SD ECDN is sometimes referred to as peer-to-peer technology, however, it encompasses much more than peer-based delivery and is thus more accurately termed a network, grid, or mesh solution.

Control Mechanisms

Regardless of how smart a particular network delivery technology may be, it cannot determine business priorities or make the right tradeoffs between competing resources on its own, all of the time.

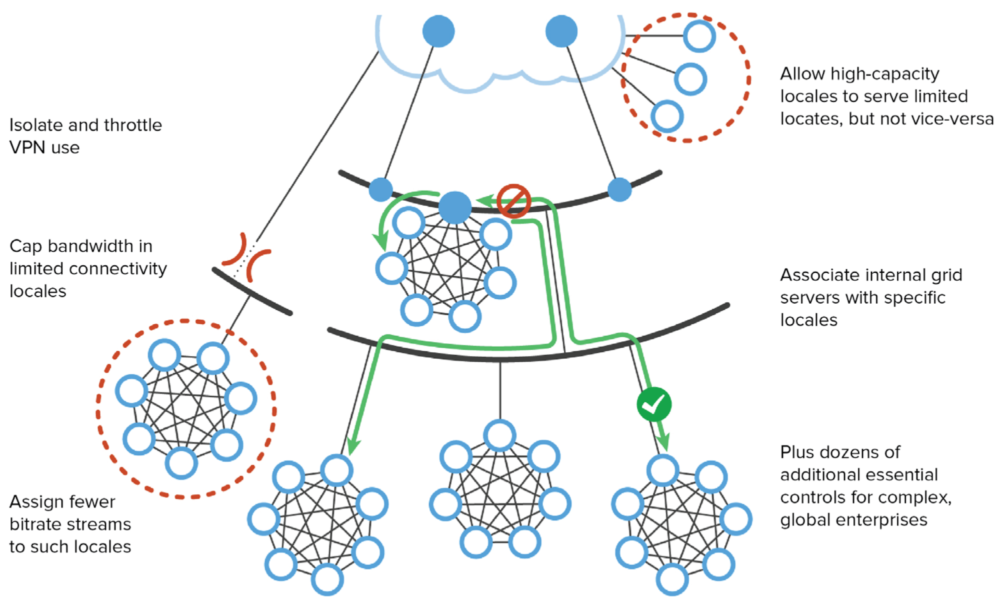
Based on our experience, the standard configuration of the Kollective SD ECDN can self-adapt to changing circumstances and constraints roughly 80-90% of the time. Control mechanisms via a configuration process are used approximately 10-20% of the time in subsets of networks where circumstances and architectures have unique requirements. The ability to rise to the challenge of the uniqueness of a given network is the most powerful attribute of the solution.

Our control mechanisms provide for extreme granularity, allowing the SD ECDN to be configured to meet those unique requirements via settings that can affect the behavior of the grid including:

- Locality settings
- Use case optimization
- Peering rules

Settings on the SD ECDN agent determine in large part how the mesh network operates. Agent configuration can be done in groups before agent deployment, or adjusted and pushed by central control as needed.

Figure 1: Control where needed



Locality Settings

Locality settings can be used to manage a wide range of attributes associated with how a given group of agents behave in the mesh. These are typically set as a range of IP addresses identified with a given locality or localities that have been selected to have a shared set of requirements.

Controllable elements with respect to localities are meant to affect:

- Use case optimization
- Peering rules
- Bandwidth

> "The ability to rise to the challenge of the uniqueness of a given network is the most powerful attribute of an enterprise video solution."

These parameters can be configured to accommodate the uniqueness of a network in, for example, remote facilities in places like South America, Eastern Europe, or smaller Asian countries. They can also be used to handle the unique needs of a manufacturing environment versus an environment designed for back-office operations.

These elements can be configured using grouping methodologies other than just by locality, they can be configured by user role, etc. Localities are a common approach and a straightforward way to describe the application of the control feature set.

Use Case Optimization

Control mechanisms can also be put in place so the mesh can adapt to different use cases such as:

- Live video
- Video on Demand (VoD)
- Pre-delivered background content

For live streaming, the mesh should be somewhat aggressive in the way it sources content from peers and the way it competes for bandwidth. Agent settings can be set accordingly for those use cases.

VoD is different, for response time is not as critical and viewer sensitivity to launch delay is reduced. For this use case, the mesh may be configured to optimize on a different outcome such as to allow a given content request to reach further out into the network to identify other endpoints that have the needed content, thus expanding the universe of good content sources. Other settings can be put in place to force the mesh to be more respectful to network traffic and more passive with regards to bandwidth contention.

Lastly, when using subscriptions or proactive content targeting to pre-deliver content, the mesh should be set to be more passive. Efforts to pre-position content are typically made with plenty of lead time and are often launched overnight from their geography of origin. In these cases, the mesh may reach as widely as possible to maximize the use of east/west traffic to deliver content while allowing little to no contention on the north/south routes.

The Kollective SD ECDN is powerful in its ability to perform specific actions as governed by the use case and ensure that the mesh attains the desired balance of performance and network load.

Peering Rules

Within the locality and use case settings as well as irrespective of them, specific peering rules can be set to govern how the agents and the mesh behave. Examples of peering rules that can be set for both the user and the serving source include:

"Setting the peering rules to include hop counts and/or latency gives control over how the software behaves while still allowing for it to optimize within the boundaries of the established rules."

- How far the agent searches to look for peer sources for a given piece of content
- How a given agent is allowed to act as a content source
- How many sources are allowed to be active as content servers to an agent

The connection between one network device to the next is called a hop. The number of such hops, known as a hop count, is a measure of the network distance between two devices and has a bearing on latency. Setting the peering rules to include hop counts and/or latency gives control over how the mesh behaves while still allowing for it to optimize within the boundaries of the established rules.

In most cases, the system uses both hop counts (number of network switching or gateway points traversed on a specific path to content peers) and latency configurations together. Network administrators can set how widely a specific group of agents is allowed to look for effective peers for the specific content item requested. The hop count method approximates whether an agent can peer from the LAN or not. Peering within 3 or 4 hops is usually within a LAN environment, so allowing higher hop counts typically means an agent will peer across a WAN segment which, in most cases, is not desirable.

The latency method consists of measuring round-trip speeds. Latency can also be a good proxy for distance in a network topology, as not all hops are created equal.

Determinations can be made as to whether a given agent is allowed to act as a content source and how many sources are allowed to be active servers to an agent. By limiting the maximum allowed bandwidth for downloading and serving, network administrators control the agent's use of network resources.

Many factors influence these bandwidth limits, and these are governed by the highest priority asset being downloaded. These factors allow the agent to aggressively stream live and on-demand content at the same bitrate.

Bandwidth limits are disabled by default so that the agent may run unfettered except when throttling is deemed necessary as a result of user activity, or to avoid network congestion.

Bandwidth Limits

The SD ECDN agent respects several bandwidth limits and different settings, such as:

Maximum download and serving bandwidth settings that limit the bandwidth used for all downloading and for serving other peers, including on-demand content and live streams.

WAN download and serving bandwidth levels that govern the maximum allowed download or serving bandwidth when downloading from, or serving a source detected as being on the other side of a wide area network link.

Absolute limits that control the maximum download and serving bandwidth the agent is allowed to use for all types of downloads and serving, including foreground downloads.

Throttle limits that control the maximum bandwidth for the agent while downloading and serving in throttled mode. The agent ramps down to these limits as quickly as possible when throttling is triggered and returns to normal bandwidth usage once throttling is no longer required. If more than one limit applies, the most restrictive limit is honored.

Device management settings that maintain the balance between increased performance and prevention of adverse impact to resources across all elements of the system. Not only is this balance in full effect on the network, tools are in place to ensure the proper balance on the end points as well. The Kollective agent monitors CPU usage to make sure it doesn't exceed a configurable, targeted rate. This includes user activity indicators such as keyboard activity, mouse movement and storage capacity.

Adapt Mechanisms

The *adapt* mechanisms of the Kollective SD ECDN make up the artificial intelligence of the mesh. This intelligence senses the intricacies of the routing and transmission environment in order to optimize network delivery efficiencies within the boundaries set during the configuration and *control* processes.

Through this adaptive and dynamic optimization of bandwidth use, SD ECDN sets itself apart from all other network delivery technologies in use today, accommodating traffic variations across the network and minimizing the network load on the WAN and Internet gateways while maximizing the quality of service delivered to viewers.

SD ECDN *adapt* mechanisms include:

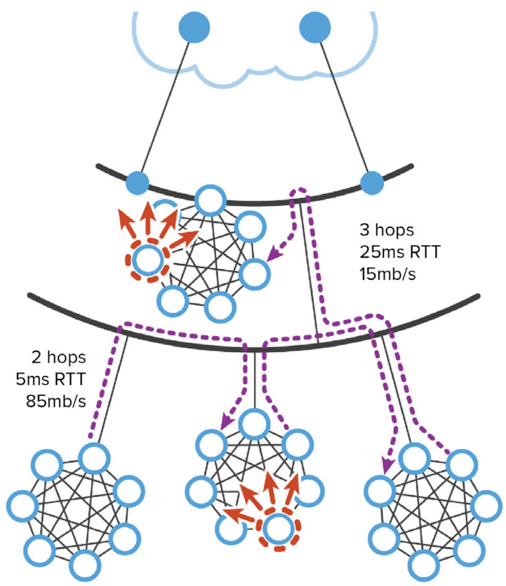
- Agent topology awareness
- Throttling subsystem
- Speed test
- Self-scalability
- Self-healing (fault tolerance)

"In combination with the control-driven attributes, the subsystem analyzes real-time data to make critical trade-offs between successful content delivery and good network citizenship."

Agent topology awareness

The agents in the mesh are constantly aware of their network surroundings. Utilizing pings and traceroutes, they gather key information about the surrounding network by monitoring the location of other agents and constantly gauging how local area links are performing. This becomes the data baseline for many of the adjustments that the mesh makes to drive efficiency during content delivery.

Figure 2: Agent Topology Awareness



Throttling subsystem

In addition to agent topology awareness, the throttling subsystem is another critical component of the mesh's ability to dynamically adapt to changes in the environment. This subsystem governs the speed at which media files are downloaded and served by the agent, taking advantage of the agent's data awareness to maximize download effectiveness.

One of the main requirements of the Kollective agent is that it acts in a well-behaved manner when sending and receiving data. The throttling subsystem attempts to achieve this goal by making optimal use of available bandwidth. In combination with the control-driven attributes, the subsystem analyzes real-time data to make critical trade-offs between successful content delivery and good network citizenship.

Speed test

Another primary adapt mechanism in the SD ECDN's arsenal is the speed test. At the time the agent launches a download activity—especially related to a live video stream—it invokes a speed test algorithm to assess the ideal bitrate for it to request given the current network environment. This real-time decision is based on many factors, but the algorithm itself can be configured to become more or less opportunistic in its selection of bitrate.

Self-scalability

As more agents join in the mesh, the distance that agents must look for high-quality peers decreases. This creates a nonintuitive situation where the more agents participating in the mesh, the more scalable the solution becomes. This effect is colloquially known as “infinite scalability.”

Self-healing or fault tolerance

As part of its self-healing adaptive power, SD ECDN adjusts for real-time changes to the network by having agents drop off that are not playing a role in the mesh for a particular download or live stream. After agents drop off, the mesh establishes connections to the next best quality peer or peers and continues uninterrupted in a matter of milliseconds.

SD ECDN’s *adapt* mechanisms are a key driver of the solution’s success in multivariate network environments. With dynamic peer quality adapting to network bandwidth throttle impacts, the system is able to thrive and deliver in any circumstance barring those that are catastrophic to the network as a whole.

Network Readiness Testing

Putting the right control and adapt mechanisms in place to ensure successful video delivery requires some advance assessment and preparation. Fortunately, Kollective SD ECDN offers the ability to test the mesh's effectiveness across the customer's network before an actual deployment using a tool called Network Readiness Test (NRT).

Exhibit 1: NRT Benefits

NRT is not a simulation, but a live test that can be executed in the background and run without the end-user being aware of its existence. NRTs are typically run by Kollective in coordination with corporate IT, who approve the test's time and range.

NRT Benefits

Kollective highly recommends performing an NRT before live deployment, for doing so helps network owners realize three principal benefits:

- Understand current network capacity limitations
- Develop a clear vision of options and problem spots
- Simulate how a video program or large file transfer will impact the network

An NRT provides essential visibility into the throughput capabilities of the network. Specifically, it can reveal network bottlenecks limiting throughput as well as infrastructure flaws that may be causing performance degradation.

A common outcome of an early NRT is the identification of places in the network that are, unknown to IT, not configured correctly for the work at hand. This pre-test allows for those changes to be made well ahead of the first large-scale, important, live event.

"An effective network readiness test simulates how a video stream or large file transfer will impact the network before the actual live event."

If there are bandwidth constraints in specific places in the network, adding more capacity throughout the rest of the network will not help. Through detailed reports, the NRT service can pinpoint offices or users that will likely experience problems in an actual live event. Kollective can then assist in resolving these issues by making changes to how the grid behaves or to how a given stream is managed within a specific locality.

Furthermore, an effective NRT can simulate how a video stream or large file transfer will impact the network. This knowledge lets network owners verify that any connectivity investments they may make will pay off when it comes time to actually holding the first major live event.

NRT Parameters

As part of the NRT, Kollective experts and corporate IT define up front the following variables for the test:

- Who to include in the test (by IP addresses)
- When they will participate (start date and stagger start time)
- How long they will participate for (duration and stagger end time)
- Participation rate (e.g., 50% turnout vs total invitees)
- Stream quality (bitrate)

The test can be run repeatedly to fine-tune the network and increase confidence.

A suite of analytics tools displays and analyzes the data to answer these three key questions:

1. How efficiently did the stream get delivered?
2. Did end user machines experience buffering, delays, etc.?
3. What was the overall delivery experience?

Poor network performance often leads to a negative user experience, which in turn may cause employees to turn against the video program and disengage more easily from future deliveries. A good readiness test like Kollective NRT ensures network professionals do not run into this problem.

Summary

The enterprise network is a living, breathing thing. The SD ECDN's ability to support the ever increasing change that lies ahead is its most critical long-term asset, and should be one of the most important considerations when looking to solve the content delivery dilemma.

The power of the Kollective SD ECDN lies in its comprehensive ability to drive the balance between network safety and optimal delivery performance. The ability of the SD ECDN mesh to self-guide and self-optimize using its adapt mechanisms, coupled with its wide-range of control features, enables network owners to master and harness the uniqueness of their network environments so content—organizations' lifeblood of communication and information—is allowed to thrive.

Why Kollective?

The largest, most successful, global companies trust Kollective Technology to power their Enterprise Live and On-Demand video delivery, serving millions of users worldwide. From its software defined enterprise content delivery network (SD ECDN) to edge related IT tools like Network Readiness Testing, and Network Analytics, Kollective drives a powerful ROI and makes the flexibility of software defined networking a reality.

Kollective

Contact Kollective

kollective.com
info@kollective.com

US +1 408 215 6400
UK +44 (0) 800 242 5602
ASIA +91 (80) 41 21 87 77