WHITE PAPER

Mellanox CDNx Reference Architecture

Contents

Executive Summary

The growth of internet traffic and new modes of web services have triggered a massive influx of data. In addition, web content is increasingly more dynamic and interactive, moving away from simple text and pictures to personalized content. High-bandwidth applications such as news feeds and streaming video are required to function seamlessly for every customer. These are some of the changes and challenges businesses face today in their quest to cater to the needs of the consumer.

Content Delivery Network (CDN) architecture must be able to respond to both the pressure of the users and the advances in technology that are making web content more complex than ever. Incompatible content from the origin servers or slow video streams deter the end-user experience, and would eventually be expensive for companies, whose success is measured by active engagement hours with users. For example, Netflix, which accounts for about 37% of internet traffic in North America, caters to the needs of around 60 million subscribers every year. Over 1 million fans tuned into Facebook Live to talk to Vin Diesel while Stephen Hawkings' AMA session, the third largest on Reddit, saw over 10,000 questions asked in just few hours. These applications put a huge demand for network bandwidth on the ISPs and the edge cache servers.

To address these challenges and prepare for the next round of technological advances, Web 2.0 enterprises need a next generation, future-proof CDN architecture such as Mellanox CDNx. Powered by Mellanox' efficient network, Mellanox CDNx can deliver a sustained throughput of up to 100Gbps while also improving the application density by offloading TCP packet processing from the CPU.

Traditional CDN Limitations

With the Cloud and Big Data boom, many online businesses today are focused on delivering personalized and dynamic content to have a meaningful relationship between brand and customer. This requires the creation of long-tail content or content that is infrequently accessed but has a long internet shelf life.

Traditional CDN infrastructure was built to deliver static content, so it cannot properly handle the acceleration of interactive web applications, constantly changing content, or personalized, localized content. Facebook, Netflix and other similar platforms have created user experiences that rely heavily on real-time interactions. . For example, Netflix provides "Movie Recommendations" in real-time for 60 million subscribers and Facebook Live allows millions of users to watch a live stream without any interruptions. Dynamic content requires infrastructure that can support millions of users and billions of interactions in a short time and on a global scale.

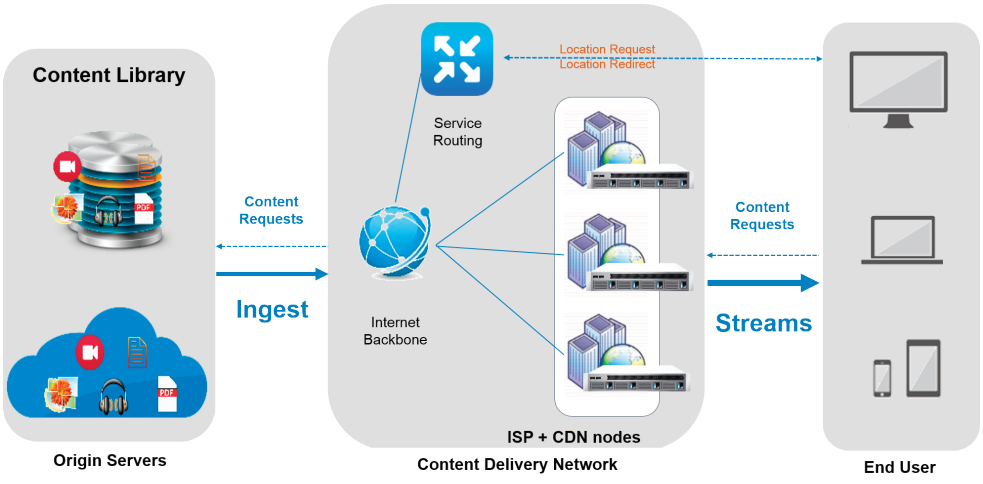

Figure 1. *CDN and Web Content Distribution*

TCP/IP Network Optimization

Typically, CDNs fetch content from the origin servers, which could be in an on-premise data center or a public cloud, cache it in various edge locations around the world and deliver it on request. This process of caching and delivering content results in the CDN acting as a one-way delivery system. This infrastructure however breaks loose when end-users begin interacting with applications, posting non-stop user generated feeds, comments, pictures and updates. The one-way delivery system must become a two-way real-time delivery system that requires different software architecture and underlying hardware infrastructure. Furthermore, newer CDN architecture like Open Connect pioneered by Netflix, are designed with the goal to get more and more bandwidth out of single box, maximizing the number of subscribers in each edge nodes. Video streaming in this newer architecture, is particularly sensitive to network delays. While increasing CPU power kept the negative effects of network at bay for a lot of applications in 1/10GbE era, the advent of 10/25 and 40/50/100GbE network has seen their return.

Though a lot of emphasis is made on the need for efficient software at content library servers, often times a good design of edge servers defines the efficiency of the CDN systems. The CDN designed for Facebook Live, for example, allows 98% of user requests to be handled by the edge servers, thereby reducing the load on the origin servers drastically. However, to cater to the demand for such dynamic HTTP workloads that are running on top of TCP/IP, network plays a crucial component in edge servers.

With the availability of faster networks at sustained speed of 100Gbps, a single edge node can theoretically handle SD video streaming requests for 512K concurrent users. But in reality, TCP packet processing in a high speed network causes significant overhead even with high-end CPUs. Packet delay and loss, misordered arrival and the unpredictable (jittery) round-trip times that are inherent to TCP/IP have drastic effects on video streaming, especially when handling a multitude of concurrent streams.

With Mellanox ConnectX-4 Ethernet adapters, CDN systems can achieve the highest throughput and application density using hardware based stateless offloads and flow steering engines, mainly :

- Packet Pacing
- Large Receive Offload (LRO)
- Receive Side Scaling (RSS)

Packet Pacing:

With bursty traffic as with TCP, packets arrive all at once. As a result, queuing delay grows linearly with load, even when the load is below capacity. With packet pacing, traffic is evenly spaced out; so there is minimal queuing until the load matches the bandwidth. Instead of transmitting packets immediately upon receipt of an acknowledgement, the sender can delay transmitting packets to spread them out, defining both the TCP window for how much to send and the rates to determine when to send.

Doing this in software for 100K concurrent streams, however, has a lot of overhead since each pacing element requires timer and context in memory. Furthermore, current mechanism in software still creates a lot of bursts, potentially causing network congestion. In contrast, by offloading the packet pacing in hardware, the NIC transmits a group of streams with same pacing, at a predefined interval that is configurable from the driver. The hardware engine supports unlimited number of rates, allowing transmission streams from different class of workloads to be processed with zero congestion. Further, this offload is designed to deliver jitter-free transmission even for the application streams that require flexible rates, as with adaptive bitrate streaming that is offered by almost all video streaming services.

Receive-Side Scaling (RSS):

Receive-Side Scaling (RSS) distributes the receiving packets across several hardware-based receive queues, allowing inbound network traffic to be processed by multiple CPUs. RSS can also be used to relieve bottlenecks in receive interrupt processing of large concurrent flows caused by the overloading a single CPU, and to reduce network latency. With RSS, incoming packets are first segregated into flows. This is determined by calculating a hash value derived from the packet header. The resulting hash value is used in a hardware lookup table that indicates to which flow or queue the packet should be directed. The hash values are also used to select a specific processor to handle the packet flow, ensuring that the packets are handled in order. ConnectX-4 adapters support around 17 million queues in hardware,

supporting a virtually unlimited number of concurrent connections, which are crucial for CDNs designed to defend DDoS attacks.

Large Receive Offload (LRO):

Large Receive Offload (LRO) is a technique used to reduce the CPU time for processing TCP packets that arrive from the network at a high rate. LRO reassembles incoming packets into larger packets to deliver them to the network stack of the system. The CPU, in turn has to process fewer packets than when LRO is disabled, which improves the utilization for networking drastically, especially in cases where end-users demand high bandwidth like a HD video. With adaptive scalable implementation, the LRO offload can decide the coalescing session dynamically between a predefined timeout and when a packet flow ends, which is determined from the packet headers.

These offloads work seamlessly in both Linux and FreeBSD systems. Currently, Mellanox is the only vendor to provide a solution for enterprise-ready FreeBSD systems, which are increasingly being considered while building CDN infrastructure for two reasons:

1. FreeBSD operating system is known to be fast and stable. Furthermore, the developer community is strong and willing to work with vendors.
2. The dynamic web applications running on the CDN appliances would be in the hands of third parties, and many web companies prefer to choose projects that use a BSD-style license rather than the GNU Public License (GPL).

Mellanox CDNx Reference Architecture

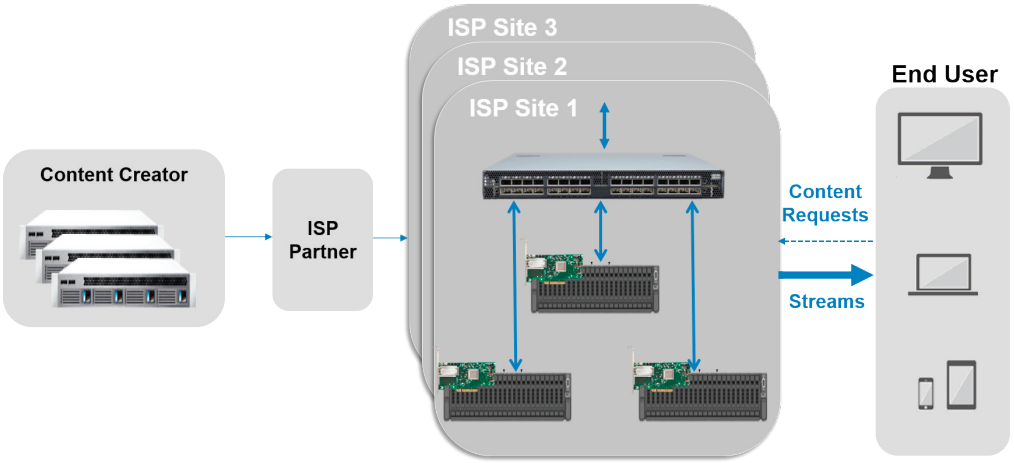

Figure 2. *Mellanox CDNx Reference Architecture*

Mellanox CDNx reference architecture provides the opportunity for ISPs to improve the user experience that their customers have for viewing dynamic web content like video streaming by localizing traffic and minimizing the delivery of traffic that is served over a transit provider. Figure 2 shows an example of Mellanox CDNx deployed in a partner ISP. Each site can deploy a single CDNx or a set of CDNx depending on specific requirements such as density of traffic in individual locations and connectivity between different metropolitan regions.

	Performance & Capacity Balanced	**Capacity optimized**	**Capacity with flash option**
Network Connectivity	• Dual port 100GbE Mellanox ConnectX-4 • 100Gbps LinkX AOC or Optic • Mellanox Spectrum SN2700	• Dual port 10GbE Mellanox ConnectX-3 Pro • 10Gbps LinkX AOC or Optic • Mellanox SX1036 switch	• Dual port 25GbE Mellanox ConnectX-4 Lx • 40Gbps LinkX AOC or Optic • Mellanox Spectrum SN2410
Server Configuration	Supermicro server with: • Intel Xeon E5-2697 • 128GB DDR4 RAM	Sanmina 4U server with: • Intel Xeon E5-2600 v3 • 512GB DDR3 RAM	Sanmina 2U server with: • Intel Xeon E5-2600 v3 • 512GB DDR3 RAM
Storage Configuration	• 4x NVMe flash up to 4.8TB	366TB storage with: • 36x 10TB hard drives • 6x 1TB SSD for caching	250TB storage with: • 24x 10TB hard drives • 10x 1TB SSD
Software & OS	• FreeBSD and NGINX	• FreeBSD and NGINX	• FreeBSD and NGINX

Table 1. *Mellanox CDNx Reference Architecture*

As shown in table 1, Mellanox CDNx solution includes the following components in a standardized configuration that scales from entry-level designs for a single appliance up to large, high-performance workloads for multiple appliances. Available reference architecture blueprints offer a choice of high-performance and high-capacity options, selected according to the specific requirement of the ISPs.

- **Performance and capacity balanced:** This configuration provides an excellent balance of computing power and storage capacity for bandwidth hungry video streaming applications. It supports 128GB of DDR4 RAM, up to 4 NVMe flash drives and network bandwidth of 100Gb/s Ethernet powered by Mellanox ConnectX-4 adapters. The next-gen Netflix Open Connect Appliances is based on the key components from this skew and is powered by 100Gb/s Ethernet using ConnectX-4 adapters and LinkX cables
- **Capacity optimized option:** This configuration developed jointly with Sanmina, provides a high capacity storage configuration for storage-intensive CDN applications. With support for up to 512TB DDR3 RAM, storage capacity of 366 TB and total network bandwidth of 20Gb/s with dual port ConnectX-3 Pro adapter
- **Capacity optimized with flash option:** This solution accelerates performance with a transparent, high-performance flash-memory perfect for latency-sensitive CDN applications. It support 512GB of DDR3 RAM, flash storage of up to 250 TB and Mellanox ConnectX-4 Lx adapters

Conclusion

Traditional CDNs are incapable of handling the dynamic workloads and demands of modern consumer applications because their architectures are outdated and their software-caching platforms are bloated and unable to keep up to the demands of real-time services. With next-gen applications, bandwidth alone is not enough. In addition to blazingly fast speed, CDN network must be predictable, in order to deliver contents steadily and reliably to millions of users simultaneously without so much as a hiccup. With Mellanox CDN platform, content providers can cater to the needs of their subscribers reliably and scale from hundreds of thousands to millions of concurrent subscribers with a total bandwidth of up to 100Gbps out of a single box, keeping a low footprint at ISPs.

Mellanox TECHNOLOGIES

350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

MLNX-31-419