WHITE PAPER

Demystifying Security for Hyperscale Data Centers

Computing at Hyperscale Falls Short Due to Network and Security Limitations

Executive Summary

Many industries have embraced hyperscale computing to address a variety of different use cases. For example, large enterprises building hybrid IT architectures to launch applications quickly require super-fast communication between assets that are spread across physical and virtual domains. Advanced research institutions, such as those performing genomic or aerospace research, require the ability to transfer massive datasets over their networks. Ecommerce companies, including high-velocity eRetail, use hyperscale architectures to manage connection bursts on major shopping holidays like Cyber Monday, and to deal with surges in activity resulting from the COVID-19 pandemic.

One problem with hyperscale data centers, however, is that they lack adequate security. Why? Because turning on security often creates a network bottleneck if security solutions struggle to keep pace, or if the IT infrastructure isn't properly segmented. With underqualified network firewalls acting as bottlenecks, many network engineering and operations leaders bypass the security altogether. However, this leaves the organization open to various types of attacks that can affect how the business performs even basic functions.

Securing hyperscale data centers requires an updated strategy and the technology to match.

Introduction

Digital innovation (DI) initiatives and business demands have changed how enterprise data centers are used and the performance metrics they must meet. Adapting to the demand for new network capabilities has driven the evolution of hyperscale data centers.

A hyperscale data center is one that scales efficiently and performs dynamically to meet escalating business demands. Hyperscale architectures are designed to meet unprecedented requirements for enormous capacity and astronomical performance. These demands can vary from industry to industry. Examples of activities that require hyperscale architectures include:

- **Large enterprises including cloud service providers.** Organizations using virtualization to create massively scalable virtual networks require large-scale network segmentations based on Virtual Extensible Local-Area Networks (VXLANs) and fast communications between services that are co-hosted on physical and virtual platforms.
- **Dynamic ecommerce including high-velocity eRetail.** Event-based connection bursts, such as shopping holidays, online tax filing, and filing for unemployment benefits, require the ability to handle massive numbers of user connections per second.[1]
- **Advanced research of pharmaceuticals, oil and gas, and aerospace.** The use of big data and machine-learning (ML) algorithms for advanced research requires the ability to transmit 40 Gbps and 100 Gbps "elephant flows."[2]
- **Stock exchanges.** Electronic trading infrastructure requires that market data be received with the lowest possible latency.[3]
- **Hyperscalers (massive, global technology companies).** High-speed data-center interconnects between cloud data centers to replicate data across disaster recovery (DR) sites require high-speed interfaces and high-throughput IPsec tunneling capabilities to achieve data privacy and confidentiality.[4]

In many cases, organizations operating in these industries have invested in the network infrastructure that they require. However, sourcing security solutions capable of meeting these needs presents a challenge since existing next-generation firewalls (NGFWs) do not meet the massive scale and performance needs of hyperscale architectures. These existing NGFWs struggle if organizations want to perform admission control on tens of millions of user connections per second, or implement anti-distributed denial-of-service (DDoS)) attacks along with essential Layer 4 firewalling. The resulting performance degradation causes many organizations to simply disable security functionality, fearing it will slow down their business and prevent them from optimizing the throughput and latency of their network infrastructure. But this is a dangerous trade-off: The organization's choice of delivering on escalating demands without adequate security controls is like rolling dice and hoping it won't be attacked.

Challenges of Hyperscale Architectures

Every deployment environment presents challenges for hyperscale security.

Pains of running massively scalable, virtualized services

Businesses need to be able to launch services in the most agile fashion possible to increase productivity and revenue. To maximize return on investment (ROI) benefits, services need to interoperate between physical and virtual assets.

By leveraging massively scalable technologies like VXLAN, customers can segment all virtualized services to realize a magnitude of scale that is not possible with VLAN. Virtualized services can scale up and down and move without significant operational overhead. These services are often required to communicate with other services located in the existing physical infrastructure. Most current solutions, however, suffer from low performance and high latency, and lack the essential Layer 4 security to track session state and admission controls that can decide who is allowed versus not allowed. They also don't offer the advanced Layer 7 security that would further detect threats, and enforce policy to achieve compliance and manage risks.

Event-based connection bursts overwhelm inelastic security

In other industries, the volume of each individual connection is not as significant as the total number of connections that an organization must be capable of processing in a very short time frame. On major shopping holidays like Black Friday and Cyber Monday, ecommerce sites experience extremely high volumes of consumer traffic over 24 hours—as much as 1.5 times more than the next-largest shopping days of the year.[5]

Similar event-based connection bursts occur during tax season, or the opening of ticket sales for major events or on holidays such as the Lunar New Year, and in online gaming environments, especially massively multiplayer games, in which hundreds of players may participate concurrently in 30-minute bursts.

In response, hyperscale architectures enable online tax filers, retailers, and game-hosting services to accept and efficiently process millions of incoming connections per second. The business case for investing in hyperscale is simple: Dropped connections or slow responses can result in lost sales and damage to brand reputation. As an example, on average, one to three seconds of delay in page load time results in a 32% increase in customers abandoning the site.[6]

Large network flows vulnerable to attack

Artificial intelligence (AI) and ML applications require enormous datasets, often reaching multiple terabytes,[7] for algorithm training and testing. Organizations in pharmaceutical, biotech, genomics, and oil/gas industries all require these large datasets for research. In order to process and analyze the data, these research institutions need to be able to transmit massive datasets efficiently over the network. However, transmitting these datasets efficiently requires network bandwidth that can reach up to 100 Gbps, or what's known as an "elephant flow."

Theoretically, research institutions should be able to leverage hyperscale network architectures built on routers and switches to deliver this bandwidth. However, these devices do not track session state and offer essential Layer 4 security. As DDoS attacks become more common, these devices are also vulnerable to attacks.

Additionally, the data they transmit over these connections is often sensitive and covered by data protection laws like the EU's General Data Protection Regulation (GDPR) or the Health Insurance Portability and Accountability Act (HIPAA). These laws mandate various access controls, which means network traffic must be routed through security technology such as firewalls and encrypt the flow of messages. However, most NGFWs cannot handle connection bandwidths of more than 10 Gbps. And this not only slows down research by a significant margin but also inhibits organizations to maximize ROI on their existing wide-area network (WAN) links that are procured to transmit data at 40 Gbps and 100 Gbps, all because a single flow through the existing NGFWs can only support 10 Gbps without crashing.

Firewall latency can lead to losses of millions

For stock market trading, competitive gaming, and similar industries, low network latency is extremely important. Even small delays in round-trip time (RTT) of network traffic can have a significant impact on profitability or performance.

As a result, organizations in the financial industry commonly invest in network infrastructure that provides extremely low latency for their data centers. Electronic trading infrastructure tolerates no more than 5 µs of latency.[8] In latency-sensitive contexts, many organizations configure their NGFWs in monitor mode, sacrificing security for network throughput.[9]

High-speed data-center interconnects require high-throughput IPsec connectivity

For cloud service providers and organizations operating content distribution networks (CDNs), the ability to replicate data across multiple regional sites is essential. The primary reason organizations use regional sites to host full copies of stored data is that they increase resiliency, decrease the latency of responses to customer requests, and lessen the load on the primary data center.

To make this possible, organizations require data-center interconnects (DCIs)—high-bandwidth links between regional sites tasked with supporting the synchronization of the network.[10] Since cloud service providers and CDNs often transmit sensitive or proprietary data, these links are often implemented as IPsec tunnels. Yet, at the same time, Layer 4 network security requires NGFWs to be capable of processing IPsec traffic with the same throughput as the network links. But, because most existing NGFWs can't achieve IPsec throughput in excess of 10 Gbps, the NGFWs securing these links can slow down the overall transfer of large amounts of data across data centers.

Conclusion

DI efforts, intended to improve efficiency and customer experience, require network infrastructure to evolve. Hyperscale data centers are designed to support massive network flows, connection bursts, and several other use cases.

While many organizations have deployed hyperscale network architecture, achieving hyperscale security is more of a challenge. Disabling NGFWs or placing them in monitoring mode, to eliminate network bottlenecks, leaves an organization open to attack and potentially out of compliance with data protection regulations. Not segmenting applications and IT infrastructure opens the opportunity for an intruder to enter and reach the core of the network after penetrating the edge. And these dangerous results compound even more if the attacks originate from internal and trusted users.

Hyperscale data centers require a completely radical approach to security solutions, one that can scale with escalating business demands. Unless a hyperscale security solution can handle massive user connections at scale, process tens of millions of connections per second, support 100 Gbps elephant flows, efficiently segment massive virtual environments, protect the enterprise edge with high-performance, essential Layer 4 security, and prevent DDoS attacks, all bets are off and tilt the advantage toward attackers aiming to launch various cyberattacks to disrupt the business, cause bad press, and eventually shut down operations.

[1] Marisa Sanfilippo, "The Best Days for Holiday Sales: A Guide for Businesses," Business News Daily, December 2, 2019.

[2] Rajiv Kohli and Nigel P. Melville, "Digital innovation: A review and synthesis," Information Systems Journal, January 29, 2018.

[3] "Deterministic Communications for Secure High-speed Performance: Fortinet Protects Connections to Electronic Trading Platforms with the Industry's Lowest Latency and Jitter Rates," Fortinet, September 23, 2019.

[4] "What is DCI?" Ciena, May 16, 2019.

[5] Marisa Sanfilippo, "The Best Days for Holiday Sales: A Guide for Businesses," Business News Daily, December 2, 2019.

[6] "Find Out How You Stack Up to New Industry Benchmarks for Mobile Page Speed," Google, March 2017.

[7] Mohammad Shaikh and Harsha Gurukar, "Machine Learning and HPC in Pharma Research and Development," Super Computing 2019, November 2019.

[8] "Deterministic Communications for Secure High-speed Performance: Fortinet Protects Connections to Electronic Trading Platforms with the Industry's Lowest Latency and Jitter Rates," Fortinet, September 23, 2019.

[9] Jason Pappalexis, "The NGFW Today: A Staple of Network Security in Spite of Challenges," NSS Labs, March 11, 2019.

[10] "What is DCI?" Ciena, May 16, 2019.

February 25, 2021 3:25 PM

wp-FA-demystifying-security-for-hyperscale-data-centers-744553-B

744553-B-0-EN