WHITE PAPER

a Hewlett Packard Enterprise company

HOW WAN OPTIMIZATION ENHANCES APPLICATION PERFORMANCE IN AN SD-WAN

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Not long ago, enterprises only deployed private-line services such as MPLS to connect branch office sites to the data center sites where critical business applications were hosted. Maintaining MPLS connections took significant time and capital because reaching all the branch office sites increased transport costs and required IT resources to manage multiple regional MPLS vendors. Since provisioning MPLS connections was often time-consuming and costly, and to minimize bandwidth costs, many enterprises deployed WAN optimization solutions. Soon, enterprises realized the benefits of implementing server virtualization technology and started to consolidate multiple data center sites. Additionally, with the advent of SaaS applications and IaaS workloads, enterprises began migrating many of their in-house applications to public cloud infrastructure such as AWS, Microsoft Azure and Google Cloud Platform (GCP). While all this was happening, the strain on the WAN increased since the traffic between branch office sites to the public clouds was first backhauled to the data center, increasing required MPLS bandwidth exponentially.

During the past five or six years, bandwidth costs have dropped dramatically for enterprises that have adopted SD-WAN solutions that enable secure, reliable utilization of lower cost broadband services. Enterprises can use a hybrid WAN comprising different transport services such as broadband, 4G/LTE, MPLS and others. SD-WAN allows enterprises to use broadband internet and 4G/LTE connections for branch-to-data center and direct-to-net connectivity while offering all the QoS advantages of a private line connection. And since these connections are relatively less expensive, adding more bandwidth is less of an issue. However, the distance between the user and the data has also increased with data center consolidation and public cloud options. Added latency in the network degrades application performance for certain applications such as file sharing. For instance, if a remote employee wants to transfer a file from a branch office site to a company folder sitting in the public cloud infrastructure it can take a considerable amount of time because of the distance between user and the data and the data receipt acknowledgement requirements of the file transfer protocol. If several employees have to make such transfers every day, the total time can easily add up to several minutes or hours. Previously, the majority of applications were hosted close to the user, so latency issues weren't as severe. Today, as most applications sit in public, private, or hybrid cloud environments, the latency factor needs to be addressed to meet and exceed application performance SLAs continuously.

A common misconception persists that SD-WAN reduces or eliminates the need for WAN optimization techniques. The reality is that SD-WAN and WAN optimization solve fundamentally different problems, and they are complementary when deployed in unison.

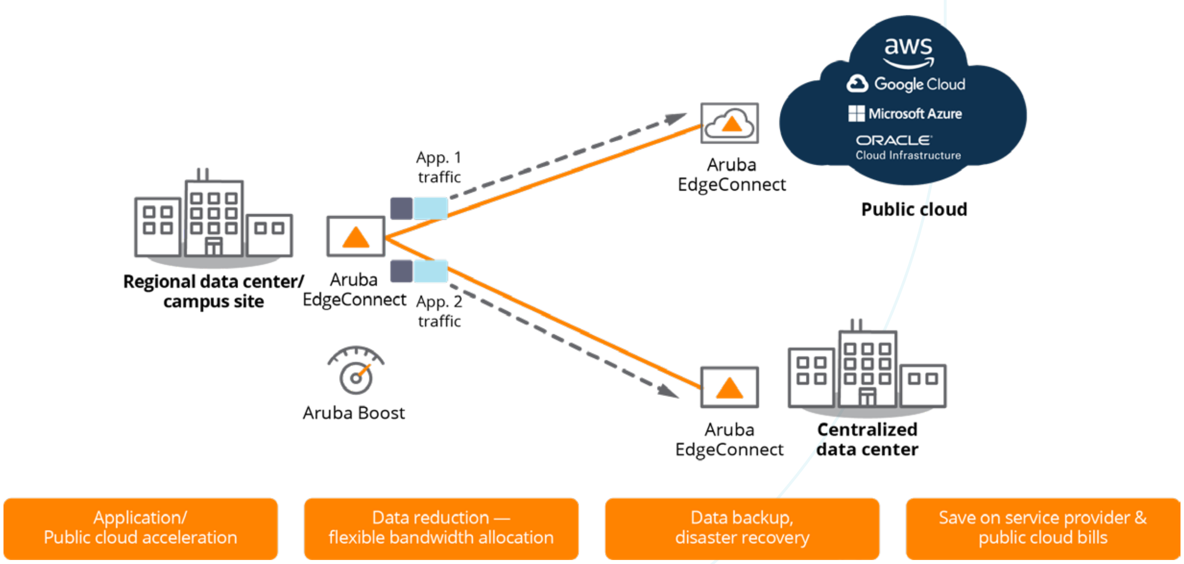

Figure 1: "One-Click" on-demand WAN Optimization with Aruba Boost accelerates applications and maximizes available bandwidth for other applications.

The time it takes for the data to travel from sender to receiver and back is referred to as network latency. As the distance between locations increases over the WAN, especially for remote international sites with low-speed transport services or where there are long backhauls, application performance degrades. This has less to do with the available bandwidth and more to do with the time it takes to send and receive data packets over distance, data receipt acknowledgements required by some protocols before sending the next segment of data and the number of times data must be re-transmitted due to packet loss. To counter such challenges, enterprises deploy WAN optimization solutions. A common misconception persists that SD-WAN reduces or eliminates the need for WAN optimization techniques. The reality is that SD-WAN and WAN optimization solve fundamentally different problems, and they are complementary when deployed in unison. Geographically distributed enterprises with locations worldwide can experience impaired application performance for critical, latency-sensitive TCP/IP applications such as transaction processing or data backup caused by excessive round-trip delays. Network latency is primarily caused by geographical distance between sites, and additional bandwidth does not change the laws of physics.

To highlight the importance of WAN optimization in today's world, let's go over three common scenario's that customers experience:

a. Many global enterprises have businesses spanning large geographical regions. To support their local business operations enterprises, open regional branch offices. These regional branch office sites communicate with the headquarter site on a regular basis to share transaction details and other business-related information. This is critical for any business to stay on track and maintain their profitability goals. However, many such branch offices are in remote areas where consistent high-speed internet connectivity is an issue. Because of this, the sharing of business information among enterprise systems is slow and that affects the speed of business. With an integrated WAN optimization solution, enterprises can confidently maintain business agility they need to succeed in today's competitive world.

b. As enterprises expand their operations it is even more critical to maintain continuous data backup and disaster recovery plans. Since many business activities are supported from regional sites, maintaining successful DRDC activities is important. But in majority of cases, the DRDC activities follow the "if it isn't broke, don't fix it" model. Because of this, the reality is that many enterprises still have legacy DRDC applications in place. To support these applications and ensure state of the art communication with the centralized data center integrated WAN optimization is needed. Without WAN optimization solution in place, conducting full and incremental backups with legacy applications becomes unmanageable and impacts application performance, causing more issues with business productivity.

c. The advent of high-speed internet and private MPLS service has eased the process of migrating large amounts of data from headquarters and regional data center site to public clouds such as AWS, Azure or GCP. However, data transfer can incur high costs, and this can be reflected in spiky cloud bills; the more data you transfer, the more you end up paying. To save money on data transfer, enterprises can deploy an integrated WAN optimization solution, that effectively reduces the amount of redundant data that needs to be transferred to the cloud via high-speed internet or private MPLS. Telco service providers also have limits on the amount of data that can travel over different connectivity options. Since, a WAN optimization solution reduces the number of bytes of data that is communicated over high-speed internet or private MPLS, customers end up saving costs on both the cloud and service provider bills.

The proliferation of SaaS and IaaS applications has put pressure on the enterprise WAN since these applications often assume LAN speeds. Although organizations can often afford to increase bandwidth as worldwide networking costs drop, latency problems cannot be overcome easily. Merely increasing the bandwidth without also mitigating the effects of high latency does not provide significant benefits to cloud-based applications. Public, private, and hybrid cloud environments all face the performance limitations inherent in today's applications and networks. For enterprises to maximize the flexibility and cost savings of the public cloud, they must overcome the latency constraints that challenge distributed IT infrastructure environments. By deploying a virtual instance of Aruba EdgeConnect SD-WAN appliance in the public cloud and enabling Aruba Boost WAN optimization, enterprises can overcome application and network latency performance problems and accelerate cloud-hosted applications and data transmission to the cloud from anywhere. As enterprises migrate more and more in-house applications to public cloud environments, they need to address the network latency limitations encountered while connecting to IaaS workloads. In cloud environments, the physical location of the company's server and the IP address or subnet might change at any time. For instance, on one day, the server can be provisioned in a region closer to the user, and on the next day it might be provisioned in a different region that is far away from the user, making it challenging for enterprises to maintain consistent application performance SLAs. With Aruba Boost, enabled on a virtual instance of the Aruba EdgeConnect SD-WAN edge platform, enterprises can leverage the WAN optimization techniques to ensure seamless public cloud integration even if the server IP address or subnets change, allowing enterprises to take full advantage of the flexibility of the cloud without any concern for application performance.

Aruba Boost accelerates data movement between data centers, branch offices, and the cloud. It uses real-time optimization techniques to overcome network quality, capacity, and distance challenges, resulting in fast and reliable access to information anywhere in the world.

Aruba Boost key WAN optimization features include:

1. Latency mitigation via TCP protocol acceleration techniques that improve application response times over distance
2. Data reduction techniques that include compression and data deduplication mechanisms to reduce the amount of data that traverses over WAN links

TCP Protocol acceleration can improve the performance of latency-sensitive applications, and data reduction techniques provide benefits when transferring large data sets, especially between geographically distributed locations. For example, accelerating replication data over distance for disaster recovery, cannot be efficiently delivered without WAN optimization. Aruba Boost is integrated with Aruba EdgeConnect and is not a separate VNF for the sole purposes of supporting WAN optimization. The flexible Aruba Boost consumption model provides WAN optimization for applications where and when it is needed. Aruba Boost is enabled in the business intent overlay configuration for classes of critical applications that benefit from it, and it is usually not enabled for applications that do not require WAN optimization such as less latency-sensitive or real-time applications like voice and video conferencing.

TCP PROTOCOL ACCELERATION TECHNIQUES

TCP is a connection-oriented and rate-limiting protocol, which is widely used in data networking and on the internet (e.g., HTTP, HTTPS, SCP/FTP and SMTP) and plays an integral role in determining overall network performance. TCP maintains a window that dictates how much traffic can flow over a connection. The TCP window is the amount of data a sender can send on a path before an acknowledgment is sent back from the receiver. TCP starts with a small window size and gradually increases the window size until an acknowledgment isn't received within a specified time period. When this occurs, TCP assumes network congestion and packet loss. The protocol responds by re-transmitting the lost data packet and decreasing the window size. Typically TCP has a 16-bit window field that is used by the receiver to inform to the sender how many bytes of data the receiver is willing to accept. In standard TCP implementation, the window field is limited to 16 bits, supporting a maximum window size of 65,535 bytes.

Aruba Boost WAN optmization software supports a window scaling feature that delivers window sizes as large as 1GB, overcoming the throughput limitation imposed by the standard 64KB TCP window size. This allows users to send more data per flow making most of the available bandwidth.

As per the Mathis algorithm:

Maximum Possible Transfer Rate (throughput) = TCP Window Size/RTT

Round-trip time (RTT) is the duration in milliseconds (ms) it takes for a data packet to go from a starting point to a destination and back again to the starting point.

From the above formula, it is clear that throughput is related to the window size. The window scaling feature allows for higher data transmission, improving network and application performance.

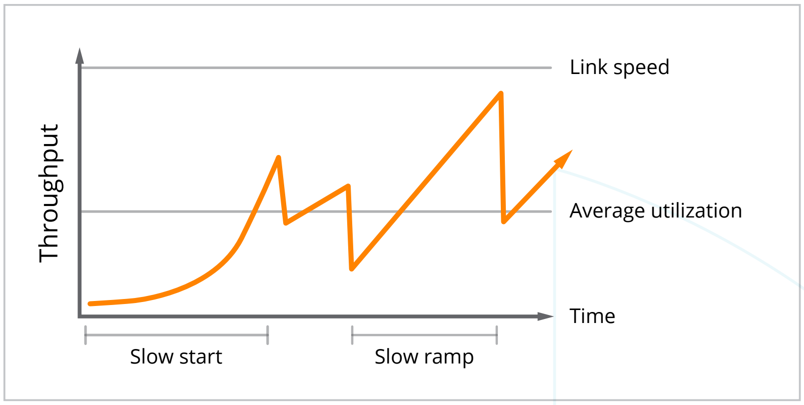

Figure 2: Throughput over a WAN link without the Aruba EdgeConnect and Aruba Boost. The average utilization of the WAN link is considerably poor since no TCP protocol acceleration techniques are applied.

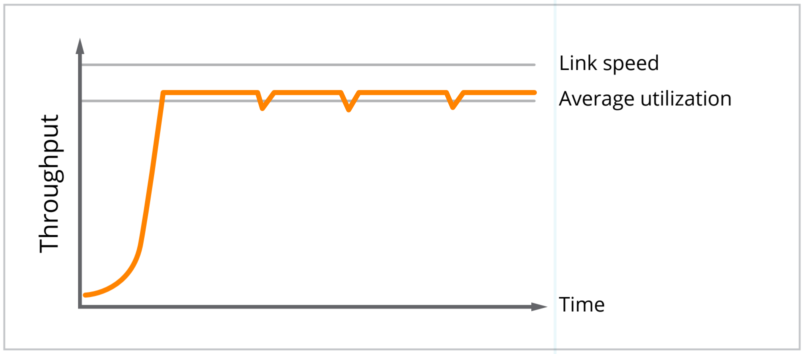

Figure 3: Throughput over a WAN link with Aruba EdgeConnect and Aruba Boost enabled. It is clear that using above mentioned TCP protocol acceleration techniques, the WAN link utilization can be dramatically improved.

Whenever TCP transmits a segment, the sender starts a timer that keeps track of how long it takes for an acknowledgment to be returned that confirms receipt of the segment at the destination. This timer is known as the retransmission timeout (RTO). As packets are sent and acknowledged, TCP adjusts its round-trip time estimate and uses this information to come up with a reasonable timeout value for packets sent. Aruba Boost employs a round-trip measurement scheme that enables more efficient RTT calculation for more accurate RTO measurements to improve network throughput. If acknowledgments return quickly, the round-trip time is short and the retransmission timeout is set to a lower value. This enables TCP to immediately retransmit data alleviating the need for a long delay. The reverse is also true. TCP does not retransmit data too quickly during times when network response time is long.

The TCP acknowledgment system can't handle multiple lost segments; it only acknowledges the last successfully received segment leading to the retransmission of data that was already received by the receiver. To counter this, Aruba Boost supports a Selective Acknowledgment (SACK) mechanism in which only the necessary packets are resent that never reached the receiver. This drastically improves network performance as fewer packets are retransmitted and results in efficient use of bandwidth. The SACK mechanism for handling multiple packet loss in a WAN environment gives a complete picture of which segments are queued at the receiver and which have not yet arrived, so senders only retransmit the missing data segments.

Aruba Boost offers High-Speed TCP, a modification to the standard TCP congestion control technique. In high-speed TCP, the TCP congestion control alters how the congestion window, which regulates the times at which the segments are sent into the network, is opened on each round trip and closed on congestion events for better performance in high-bandwidth, high-latency environments. The conventional TCP congestion control algorithm is ineffective in networks where window sizes may become very large. In high-bandwidth and high-latency networks, following a congestion event, it may take an excessively long time for the congestion window size to recover. This leads to slow network convergence properties and poor bandwidth utilization across WAN links.

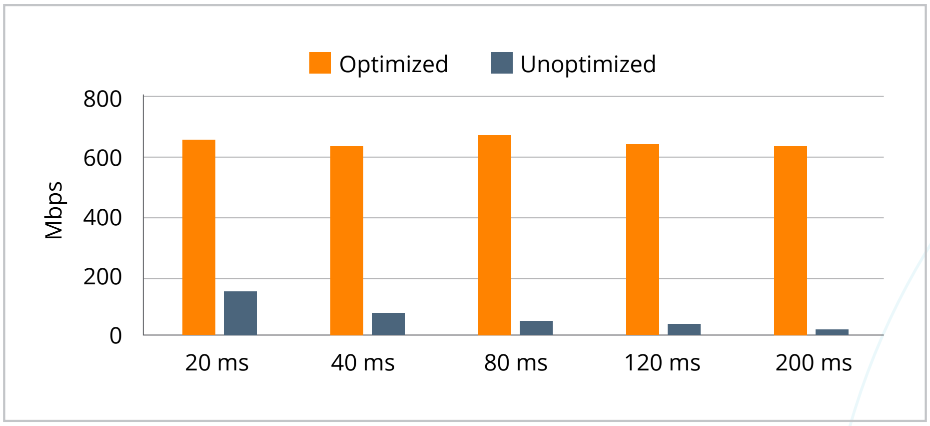

Figure 4: TCP protocol acceleration techniques accelerate applications over distance overcoming the effects of latency.

In high-speed TCP, when the window size is small, the algorithm behaves exactly like ordinary TCP. But when the window size is large, it increases the window by a larger amount and decreases it by a smaller amount. The effect of these changes is that TCP achieves much higher throughput over high-bandwidth, high-latency networks.

DATA REDUCTION TECHNIQUES

Aruba Boost WAN optimization delivers the technology needed to improve WAN performance and offer end users a LAN-like experience. Data reduction techniques allow enterprises to achieve superior application performance over a hybrid WAN comprising MPLS, internet and 4G/5G/LTE connections. The data reduction techniques include data compression and deduplication that eliminate the transmission of any redundant data, thereby improving WAN performance as fewer data packets traverse the WAN at any given time. Additionally, data reduction frees up bandwidth for other applications enabling customers to use bandwidth more efficiently and potentially save costs by deferring the need to increase bandwidth. For **data compression**, Aruba Boost incorporates a variety of compression algorithms for optimum "first pass" gains. The data compression techniques are applied both for the payload and the IP header to achieve optimal compression. Data compression techniques work on repetitive traffic and across various protocols. Payload compression uses algorithms to identify relatively short byte sequences that are repeated frequently. These are then replaced with shorter segments of code to reduce the size of transmitted data. Additionally, IP header compression is the process of compressing excess protocol headers before transmitting them on a link and decompressing them to their original state at the other end.

Data deduplication technology improves bandwidth efficiency through advanced fingerprinting algorithms that examine all incoming and outgoing WAN traffic.

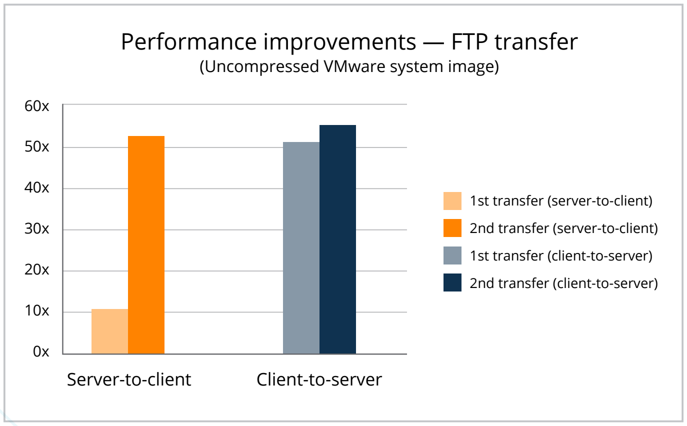

Figure 5: Network memory is bidirectional, providing dramatic performance improvements in both directions of a file transfer.

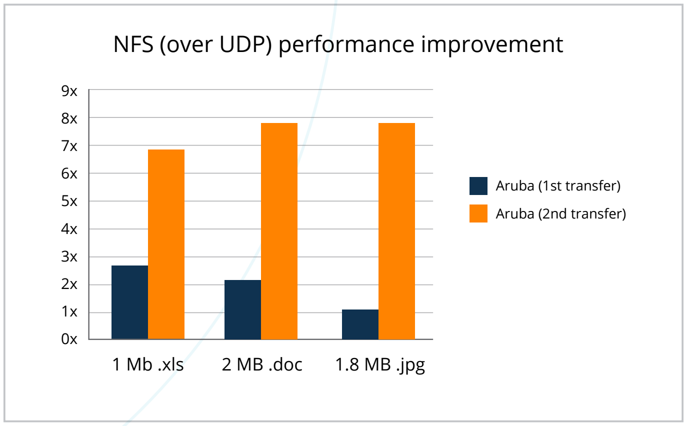

Figure 6: Network memory works at the network layer, providing significant performance improvements across all TCP and UDP applications, including NFS.

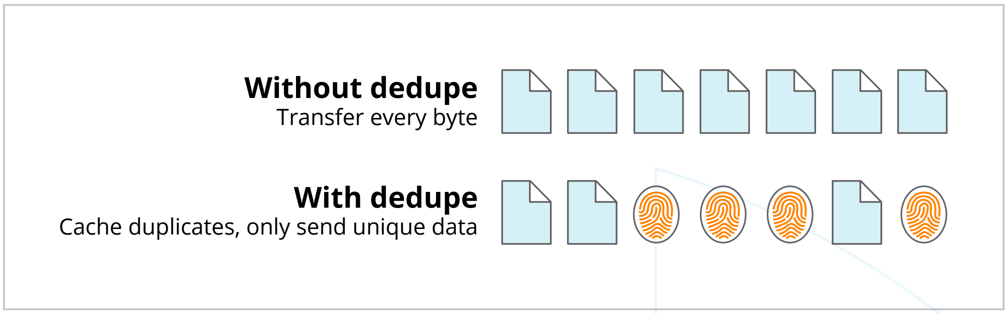

Figure 7: Data deduplication eliminates overhead of redundant packets traversing the WAN.

Network memory localizes information and transmits only modifications between Aruba Boost-enabled Aruba EdgeConnect devices. Since network memory is bidirectional, pattern recognition can be performed across flows, and patterns learned in one direction can be utilized in the opposite direction without having to be re-learned. Network memory operates at the network layer of the OSI stack. As a result, unlike other WAN optimization solutions, Aruba Boost is able to accelerate the performance of all file sharing applications including those that use UDP, such as NFS.

For data deduplication, Aruba Boost uses device peering with tunnels as a communication method between Aruba EdgeConnect appliances. This also allows real-time Path Conditioning features such as Forward Error Correction (FEC) and Packet Order Correction (POC), which are critical functions for overcoming adverse effects of dropped and out-of-order packets inherent when using broadband internet or 4G/5G/LTE connections and even satellite or microwave connections.

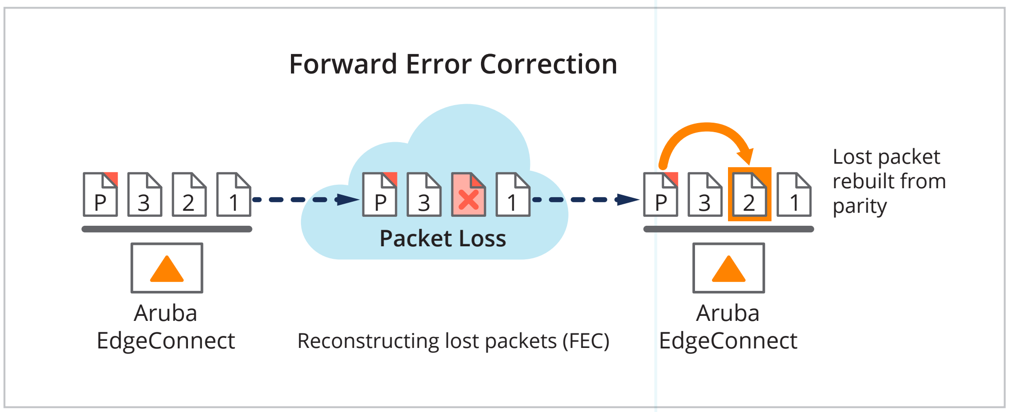

Figure 8: Using Forward Error Correction, packets that are lost in transmit across the WAN are rebuilt using parity packets. Aruba EdgeConnect dynamically adjusts FEC in response to changing link conditions to minimize overhead.

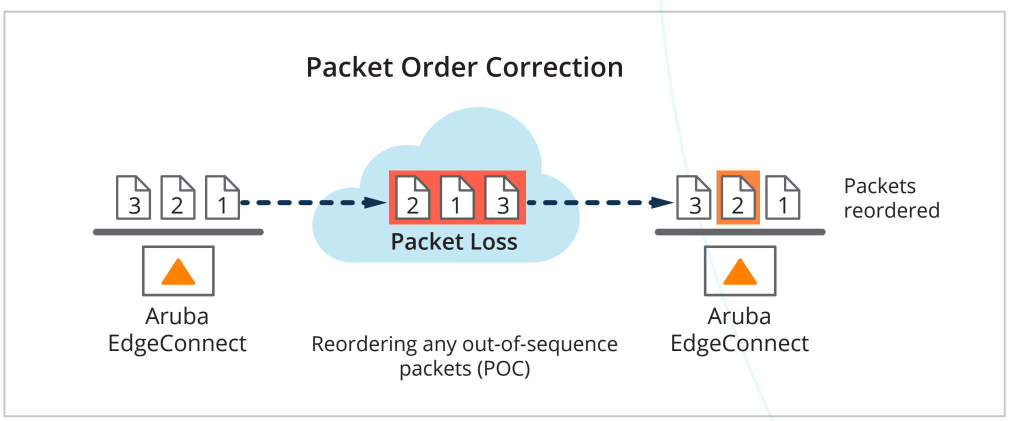

Figure 9: With Packet Order Correction, packets delivered out-of-order across the WAN are reordered into the correct order.

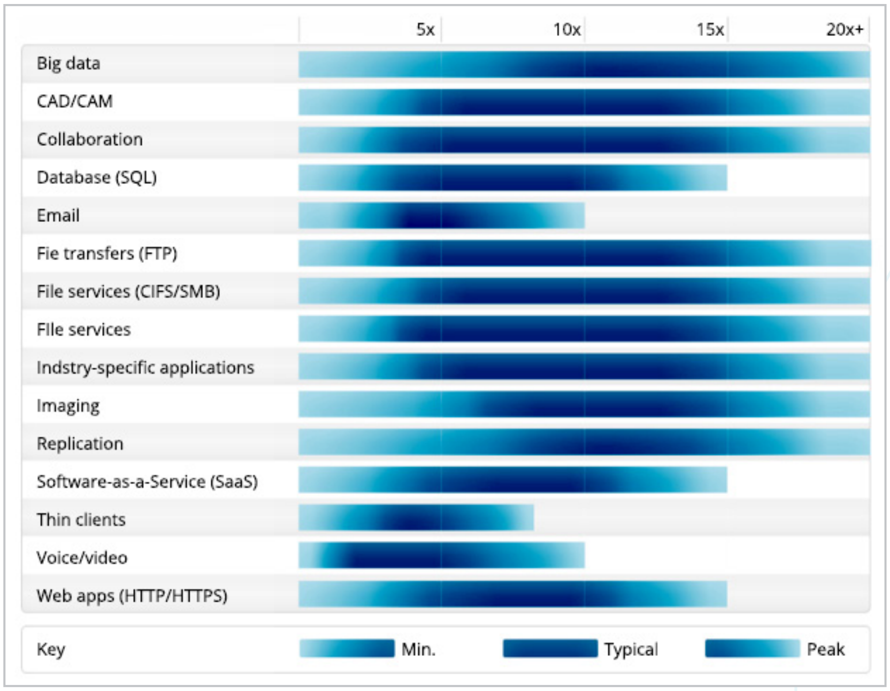

Table 1 shows application performance gains achieved by enterprises through latency mitigation with Integrated WAN optimization solution from Aruba.

In enterprise networks, most applications include Secure Sockets Layer (SSL) encryption as a default. With Aruba Boost enabled on Aruba EdgeConnect devices enterprises can enable data reduction techniques for SSL traffic by decrypting the SSL session, applying compression techniques, deduplicating the data, and then re-encrypting the data for transmission across the WAN. This process improves an applications' performance by reducing the amount of data that needs to travel over enterprise WAN.

Aruba Boost WAN optimization techniques improve the performance of data center-hosted applications. Today, as enterprises continue to migrate applications to SaaS and IaaS workloads, they realize the network and application performance challenges associated with the migration. Customers facing these challenges can deploy a virtual instance of Aruba EdgeConnect in the public cloud and benefit from Aruba performance features such as tunnel bonding, path conditioning, and Aruba Boost. Aruba Boost improves the application performance by minimizing the impact of latency over the first mile between the branch office site and the cloud using TCP protocol acceleration and data reduction techniques. Providing inherent resiliency ensures applications perform to service levels, and users remain productive regardless of failures of underlying network transport services.

The WAN is no longer just a pipe that connects point A to point B; it has become a strategic asset that enables employees to be more productive, businesses to be more competitive, respond more quickly to customer and industry demands, and serve as a foundation for whatever innovations might evolve next. The challenge of delivering applications to users at branch offices and data center sites with high performance has not gone away. The rise of cloud networking, SaaS applications, and IaaS workloads has made the issue more complicated. Aruba EdgeConnect with Aruba Boost can improve the connectivity to cloud-based applications, accelerate performance and optimize bandwidth utilization between branch offices, enterprise data centers, and public cloud environments.

The three scenario's listed previously in this paper highlight the significance of WAN optimization in an SD-WAN world. With an integrated WAN optimization solution, enterprises can confidently maintain business agility they need to succeed in today's competitive world. As enterprises move, more and more of their data to public clouds, seamless connectivity to public clouds and in-between them is critical for businesses to run smoothly and profitably. By spinning up a virtual instance of an Aruba EdgeConnect SD-WAN appliance in the public cloud and enabling Aruba Boost WAN optimization, enterprises can overcome application and network latency performance problems and accelerate cloud-hosted applications and data transmission to the cloud from anywhere. Additionally, routine data backup and recovery activities involving legacy applications become manageable ensuring optimal application performance while minimizing negative impact on business. Furthermore, cloud data transfer bills and telco service provider bills are reduced drastically through the use of WAN optimization.

REFERENCES:

- SD-WAN experts Blog: Why your maximum throughput is less than your bandwidth
- Scalable TCP: Improving Performance in Highspeed Wide Area Networks by Tom Kelly
- TCP Congestion Control research paper

WP_WAN-Optimization_EM_031821 a00110933enw

Contact Us Share