

Whitepaper

Controlling the Network Edge to Accommodate Increasing Demand

February 2007

Introduction

A common trend in today's distributed work environment is to centralize applications and the data previously hosted in individual branch offices. The driving factor for this centralization is the need to simplify management of the application infrastructure, which in turn helps to reduce cost and increase overall control and performance. While there are clear benefits with such a model, this centralization also comes with its share of challenges for those responsible for the application and network infrastructure.

With the number of users outside of the enterprise expanding through a mix of mobile workers, branch office employees, partners and customers, the dependency on network-based applications increases as well. Along with this increase comes the importance of ensuring that those applications are available around the clock. In addition to the application infrastructure, the edge of the network has now become an important lifeline as the mechanism to connect those on the outside with the critical services now found within. Should this lifeline fail, employees, partners and customers could be left without access to critical information and services, limiting their ability to perform transactions which often times translates directly to a loss of revenue or decrease in customer satisfaction.

Slow applications can also have adverse effects on a company's bottom line and satisfaction rating. With a decrease in response times an application may slow to the point of being unusable, hindering the ability for employees to do their job and causing customers to seek other solutions. While all of this has direct effects on the bottom line, it becomes increasingly important to properly manage the network edge as part of any application consolidation effort. If not properly managed, a project which has set out to save an organization money may end up costing more than anticipated.

Ensuring Application Availability

Availability of services at the edge of the network is essential to guaranteeing availability of critical applications within the data center. Should a WAN link fail, the obvious result is the loss of connectivity between the outside world and anything behind the edge of the network - leaving remote and mobile workers without access. The same may be true should any of the services at the edge fail which may include security, caching and web or application servers.

To ensure availability of services, traffic load can be distributed, or load balanced, across 2 or more entities providing a service (i.e. a WAN link, firewall, web or application server, etc...). With no single point of failure, the chances of losing the connection to the applications are diminished. Load balanced architectures will also help to simplify tasks such as upgrades and maintenance tasks - since one device in the pool can be taken out of service for a period of time to perform the upgrade while the other devices continue to serve traffic. In addition, this type of architecture will help organizations accommodate the need for increased capacity with the ability to add additional devices to a pool over time as demand requires.

Accommodating Growth

As users shift from accessing applications locally in branch offices to a model where services are centralized, traffic across the enterprise network into the data center will naturally increase. This increase will not only impact the applications and the data center, but will also impact the supporting network infrastructure, which includes the network edge. With this increase in traffic, business critical applications will be vying with non-business traffic for limited bandwidth across the network and the WAN.

While adding WAN bandwidth can address some issues associated with new capacity demand, this approach can come with an undesirable, reoccurring expense. As a proof point, adding an additional T1 link to address increased bandwidth requirements can add an average of \$900 per month* to operational expenditures.

Another approach to addressing the increase in bandwidth requirements is to deploy lower cost, business grade DSL lines. This approach can provide the same level of capacity for a fraction of the cost (average \$150 per month). While in some cases a DSL line may not be as robust as a T1 link, this type of service can be implemented with Link Load Balancing technologies to provide cost effective, scalable and highly available services. In addition to better performance and redundancy, Link Load Balanced connections

can also be used to accommodate traffic differentiation where critical business traffic is routed over high-grade connections, which guarantee latency and up time while non-business traffic is passed through lower cost non-guaranteed ISP links.

**based on average pricing for broadband services in North America.*

Controlling Bandwidth Usage

Since traffic across the network and WAN can be a mix of critical business applications and casual web traffic, it becomes increasingly important for network administrators to understand which individual applications and protocols comprise that mix. With the increase of users depending on the network to perform daily functions as well as for casual surfing, it is not at all uncommon for this mix of traffic to create a situation where critical applications are contending for limited resources. In some cases, where network congestion is present, the resulting environment may be one where critical business data is dropped and casual data is let through. Mechanisms that enable the right level of visibility into the network can help administrators better understand the mix of traffic present in the network which can be used to properly prioritize traffic.

With visibility into the types of applications that are competing for resources, network administrators are better positioned to react to network events such as traffic spikes and plan for future growth to minimize congestion. With technology such as Quality of Service (QOS) and bandwidth rate shaping, network administrators can control the amount of resources that a specific application can take - allowing for the prioritization of critical applications over non-business traffic. In addition, with this knowledge, network administrators can control which egress links traffic should take, providing the ability to send business critical applications over high quality links while sending non-business critical traffic over lower cost, best effort connections.

The ability to effectively classify applications reaches beyond the basic port level with the need to identify applications that use dynamic or user configurable ports. In many cases these types of applications, which include P2P, streaming video and other rich media, not only fall into the category of non-business critical but can also consume the bulk of bandwidth available to an organization. Solutions that can identify and control applications based on a range of criteria can properly classify and prioritize all types of traffic to ensure that they are correctly directed through the network.

User Level Visibility

Knowing which applications are being used throughout the network provides a portion of the information needed to effectively manage the network resources. Another key piece of information is knowing who the user is that is using those applications and consuming bandwidth. Having the ability to drill down to the user level will not only help identify the need for additional bandwidth resources, but it will also help identify the persons responsible for any misuse of corporate resources – allowing organizations to quickly enforce their corporate usage policies and control access to corporate resources.

Associating user identity with application traffic can greatly help network administrators instantly identify the top users of specific applications. With this information, network teams can address any anomalies at the source to eliminate the need to change infrastructure to react to increased traffic loads or misuse. In the event that infrastructure change is required, this information can also help administrators proactively plan for additional capacity or services - targeting these upgrades at specific groups in the organization.

Associating user information with application usage can also integrate with and help to simplify compliance tasks. With this information, the correlation between an application transaction and the user behind that transaction can easily be made and eliminate the need to manually search through log files for this information. With the ability to quickly respond to questions dealing with specific application usage, organizations can save time and money when working with auditors and compliance teams.

Protecting the Edge

In addition to being a life line to outside users, the edge of the network is the first line of defense for the internal network and the services within the data center. As network speeds and the demand for access to services within the data center increase so too must the security infrastructure. A scaleable, multi-layered approach to security will provide flexibility as well as accommodate requirements for capacity into the future.

A multi-layered security architecture provides protection against many types of attacks while ensuring that there is no single point of failure. Spreading this functionality across several types of devices also allows for organizations to protect against attacks that vary in size and type.

Another challenge is the ability to scale the existing security infrastructure that organizations have in place. There are two main options to address this requirement:

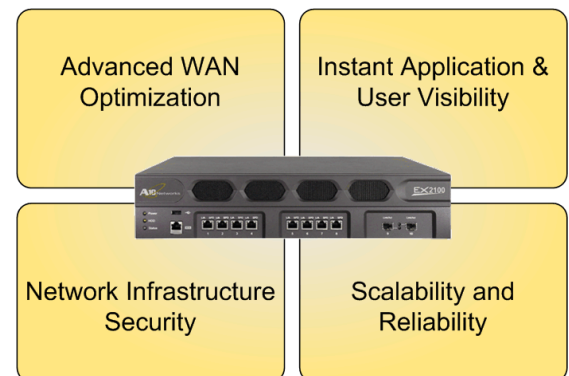
- 1) Deploy new, higher capacity devices – While this approach may provide organizations with ample room for growth it often translates into a “forklift upgrade” to completely replace existing solutions with new hardware. In addition this will likely include the overhead of coming up to speed on the new tools so that they can effectively be deployed and managed.
- 2) Deploy additional devices of the same kind – Using load balancing technology, additional devices of the same type can be deployed as needed to add incremental capacity. This method is not only easily scaleable but also protects the edge from any single point of failure. In addition, the learning curve is eliminated since more of the same device is deployed over time.

The ability to distribute load across devices at the edge not only requires performance but also requires intelligence to ensure that connections from the same session traverse the same firewall both into and out of the organization. This is a capability that is not inherent in every load balancing product and requires a combination of traffic distribution, health monitoring and session persistence capabilities to properly fail over and maintain user connections.

A10 Network Secure WAN Acceleration Solution

The EX Series from A10 Networks helps organizations better manage WAN connections and the services which are critical at the edge of the network. This not only ensures connectivity to outside users but also helps to ensure availability of critical applications in the data center. In addition, the EX series helps to prioritize traffic to ensure that important applications have priority through the network and across the WAN. With the EX Series, A10 Networks has combined the critical functions for managing the edge of the network into a single, purpose built, high performance network appliance allowing customers to simplify deployments and minimize the overhead of both management and latency.

Advanced WAN Optimization – The EX Series helps organizations ensure peak performance of services through the network edge and across the WAN. With the ability to distribute both incoming and outgoing traffic, the EX can not only help organization scale critical services but also serves the essential task of ensuring uptime for these services. In addition, the EX provides maximum application performance with technology to prioritize, shape and compress traffic traversing the network edge. The EX can identify application traffic beyond simple port levels allowing for a wide range of application support and greater flexibility. With this level of control, the EX Series helps to enhance the delivery of critical traffic which improves application response times and application usability across the WAN.



Instant Application & User Visibility - The EX Series delivers a unique solution giving IT groups visibility into protocols running in the network as well as the users of those protocols. This unique feature provides network administrators with a better understanding of who is using the resources and how they are being used. With the ability to identify such things as top users for specific applications network administrators are better positioned to react to and adapt to evolving network conditions.

Type	Date/Time	User Name	App User Name	Source IP	Destination IP
yim/logoff	Oct 27 17:13:32	cparker	yardbird829	192.168.43.161	216.155.193.141
yim/logon	Oct 27 17:13:29	mtyner	bones1211	192.168.43.180	216.155.193.142
yim/logoff	Oct 27 14:01:40	srollins	sonny0907	192.168.43.165	216.155.193.174
yim/logon	Oct 27 13:59:07	srollins	sonny0907	192.168.43.165	216.155.193.174
msnim/logoff	Oct 27 10:01:26	dcherry	donc_msn1936	192.168.43.181	192.168.230.81
yim/logoff	Oct 27 10:01:13	dcherry	donc_yim1936	192.168.43.181	216.155.193.181
yim/logon	Oct 27 09:27:10	cparker	yardbird829	192.168.43.161	216.155.193.141
aim/logon	Oct 27 09:04:57	gevans	arranger1912	192.168.43.151	205.188.179.233
yim/logoff	Oct 27 08:48:09	tmonk	mrmonk	192.168.43.174	216.155.193.149
yim/logon	Oct 27 08:44:42	tmonk	mrmonk	192.168.43.174	216.155.193.149

Example of EX's Identity-Based Reporting Capabilities for Instant Messaging Users

Network Infrastructure Security – The EX series provides a first level of defense for the network edge with the ability to identify and mitigate against a long list of possible network-level DDoS and protocol anomaly attacks. This capability ensures that critical edge services remain available under any condition. In addition, the EX Series includes a rich set of traffic distribution capabilities which include the ability to distribute traffic across a pool of firewalls or other security devices. With this capability, the EX Series helps organizations easily and incrementally scale existing security infrastructures to meet growing demands.

Scalability & Reliability – The EX Series is built upon a highly scaleable Symmetric Multiprocessing (SMP) architecture allowing for the support of multiple features enabled simultaneously without sacrificing overall device performance. On top of this foundation, the EX combines critical features required to successfully manage the network edge and related services allowing for the EX solution to scale as traffic demand on the edge grows and helping to ensure top WAN performance. In addition, the EX Series delivers solution redundancy at multiple levels with features within a single chassis and across multiple chassis to ensure uninterrupted service.

For more information on the EX Series and other A10 Networks products to help you Accelerate, Optimize and Secure your network, please visit: <http://www.a10networks.com>

Contact Information

Corporate Headquarters

A10 Networks, Inc.
2309 Bering Drive
San Jose, CA 95131
USA

Website

<http://www.a10networks.com>

A10 Sales

N. America: +1-888-A10-6363
sales@a10networks.com

International: +1-408-325-8616
sales@a10networks.com

China: +86 10 5172-6675
china_sales@a10networks.com

APAC: +886-2-3322-5882
apac_sales@a10networks.com