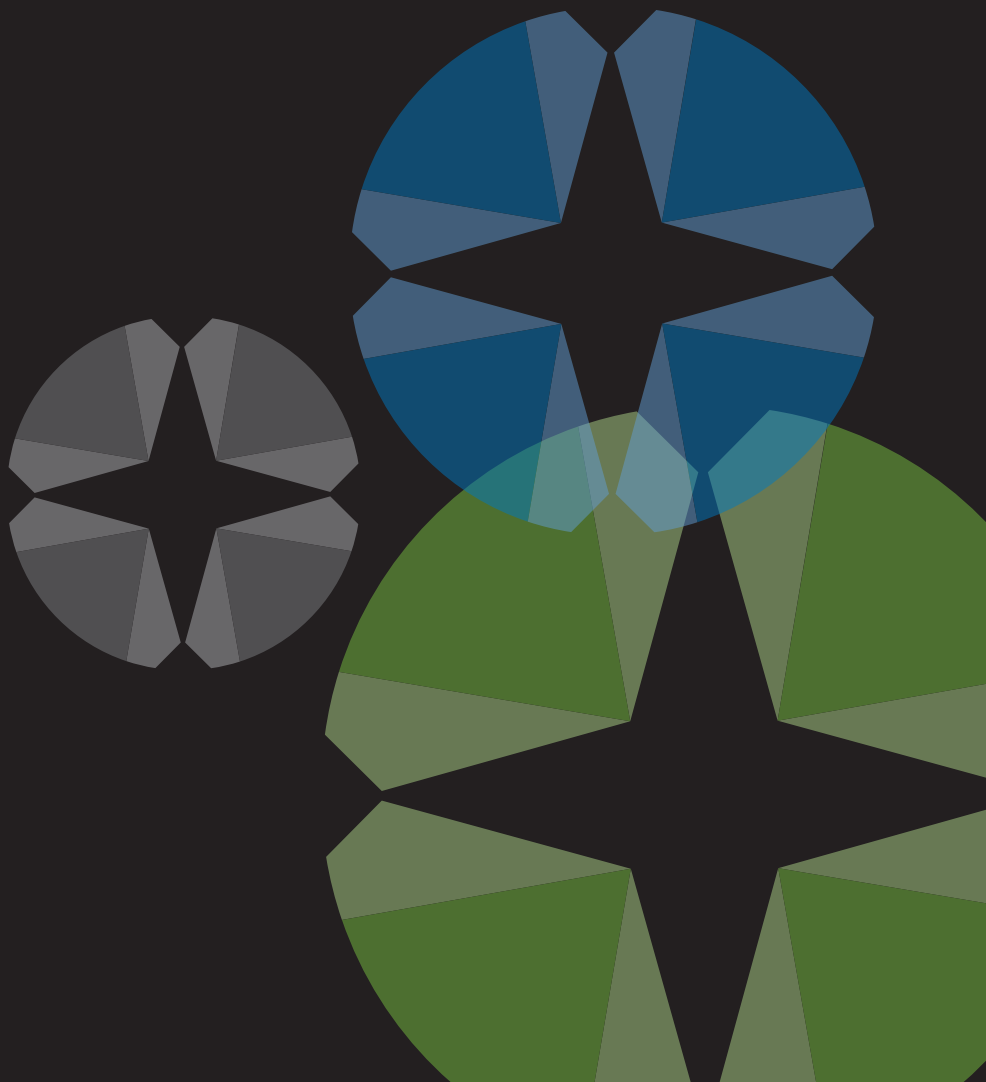
exinda

Getting the Most Out of Your Existing Network

A Practical Guide to Traffic Shaping



exinda

Getting the Most Out of Your Existing Network

A Practical Guide to Traffic Shaping

Executive Summary

As organizations face increased competition, uncertain economic times, and reduced IT budgets, they must evaluate all aspects of their IT infrastructure. In some cases, individual users or applications get a disproportionate amount of the network resources and as a result critical business applications may slow and be negatively impacted by non-critical or recreational network traffic. Without effective traffic shaping, an organization will not be able to manage and measure the behavior of their applications or the traffic on their networks. This results in decreased user productivity and satisfaction as well as potential increases in costs due to misappropriated network resources. In addition, organizations often face added costs in bandwidth upgrades purchased in an effort to improve the performance of slow applications or to increase user satisfaction or productivity.

In this paper we will explain traffic shaping, the technologies that are required to implement effective traffic shaping, and how the process an organization can follow in identifying and fulfilling their need for traffic shaping.

What is Traffic Shaping?

Traffic shaping (also known as "packet shaping") is the control of network traffic in order to optimize or improve application performance, reduce network latency, and increase usable bandwidth on a wide area network. Traffic shaping provides a means to control the volume of traffic being sent into a network (bandwidth throttling), and the maximum rate at which the traffic is sent (rate limiting). Traffic shaping or traffic management involves controlling the flow of traffic on the network so that the highest priority applications or users always have adequate bandwidth. There are a variety of techniques used to gain visibility and shape network traffic that we will review later in the document.

The Importance of Traffic Shaping

Traffic shaping allows an organization to realize the full potential of their wide area network resources and enables them to proactively manage and plan for future needs. In many cases, an organization will find that implementing traffic shaping will allow them to forego added costs or reduce existing costs by managing network usage while improving application performance, employee productivity and the end user experience.

The Challenge: High Speed LAN's, Low Speed WAN's

Data networks that span large distances are known as wide area networks or WAN's. They have much different requirements and characteristics than local area networks, or LAN's, that span across shorter distances like campuses and buildings. WAN's operate at much lower bandwidth speeds than LAN's with a typical rate between 1.5 Mbps to 155 Mbps compared to 100 Mbps or 1 Gbps on a LAN. In addition, since the WAN network spans much larger physical areas, between states or countries, across oceans or continents, the amount of time it takes for the data to travel across the network is much longer; this is referred to as higher latency. In addition to having higher latency and lower bandwidth speeds, the costs associated with procuring, maintaining, and managing WAN's are far greater than those with LAN's.

With the push to server consolidation many applications that have been designed and implemented for delivery over local area networks are now being delivered over WAN's. This significantly increases network traffic. In addition, routers are the primary interconnecting devices on wide area networks and their capabilities do not support the requirements of today's networks, users, and applications. Routers have limited traffic shaping capability which leads to indiscriminate packet drops. Also, when dealing with congestion, dropped packets and retransmissions are the norm because the router was not designed to deliver application performance, it was designed for packet forwarding. As such, routers offer no fair allocation of bandwidth between sessions or users and provide ineffective mechanisms to guarantee delivery of critical applications. All these lead to poor performance of applications over a WAN.

Traffic Shaping Technologies

Many underlying technologies integrated together are able to deliver the core components needed for effective traffic shaping. Network visibility is the foundation and the first requirement for traffic shaping. If you cannot identify all the traffic on a network, you cannot effectively manage that network. Once traffic has been identified it then needs to be shaped based on policies of the organization. These policies deliver the appropriate network resources aligned with the organizations business objectives. Finally, the traffic must be optimized reducing the amount of network resources are used and improving the performance of the applications and user experience.

Network Visibility

Network visibility techniques provide a deep understanding of application, user, and network behavior. As applications within the network change and the network evolves, network visibility allows for a better understanding of application performance and user expectations. Visibility is the cornerstone of network traffic shaping. Without detailed visibility, an IT organization will struggle to gain an detailed understanding of their existing network environment.

Deep Packet Inspection

One element of network visibility is provided using deep packet inspection technology. Deep Packet Inspection (DPI) provides visibility and application awareness by performing analysis of the content on the network. Similar to how an x-ray can identify the contents of a package or a fingerprint identifies a person's identify, DPI inspects the packet headers and protocol structures as well as the actual payload of the message. DPI will classify the traffic based on a database of signatures that includes information extracted from the data part of a packet. This allows for finer control compared with other classifications based only on layer 4 of the OSI model. Deep packet inspection inspects the complete communication including all layers of the OSI model including layer 7. This provides the ability to analyze network usage and optimize network performance.

Heuristic Inspection

Heuristic inspection allows visibility and control of a new breed of advanced applications such as Bit Torrent and Skype that use full encryption techniques to mimic other applications. Heuristic inspection is an additional requirement for complete traffic classification in today's networks. A few years ago, the level of sophistication of many protocols and applications like BitTorrent Peer-to-Peer increased to a level where deep packet inspection was ineffective. These applications have always been difficult to identify but deep packet inspection was capable of doing so. With the increase in sophistication these applications and protocols now completely encrypt the communications which scrambles the packet payload beyond recognition. There is no use of SSL certificates or other means to identify the traffic so heuristic based inspection was developed to classify the traffic based on behavior by correlating a number of attributes of the traffic such as packet size, packet frequency, connection frequency, and more. Essentially heuristics based classification relies on a behavioral based signature to identify the application.

Application and Protocol Signatures

The Exinda Web-based user interface provides a real-time dashboard with single pane of glass visibility into all network activity. Network administrators simply log in to the management interface to obtain a live view of network activity.

At a glance, administrators can identify performance issues as they are happening and rapidly troubleshoot the network. Alerts, warning and email notifications automatically inform administrators of configuration errors, NIC drops and transmit/receive errors.

Application Response Time Measurements

Application response time(ARM) measurements quantify application performance from the end-user's perspective. Application response time measurement detects how long end users are waiting for their applications to respond. Wide area networking is no longer about connecting locations to a WAN cloud and ensuring connectivity, but rather it is about understanding networking applications so they perform to meet the needs of the individual end users as well as those of the entire organization. ARM helps to pinpoint whether a performance problem is network or server related. Equipped with this information, network managers can fine tune traffic shaping policies to control application response times. These measurements are an integral part of proactive traffic shaping. Without monitoring application response times, there is no clear measurement to know if application response times are improving.

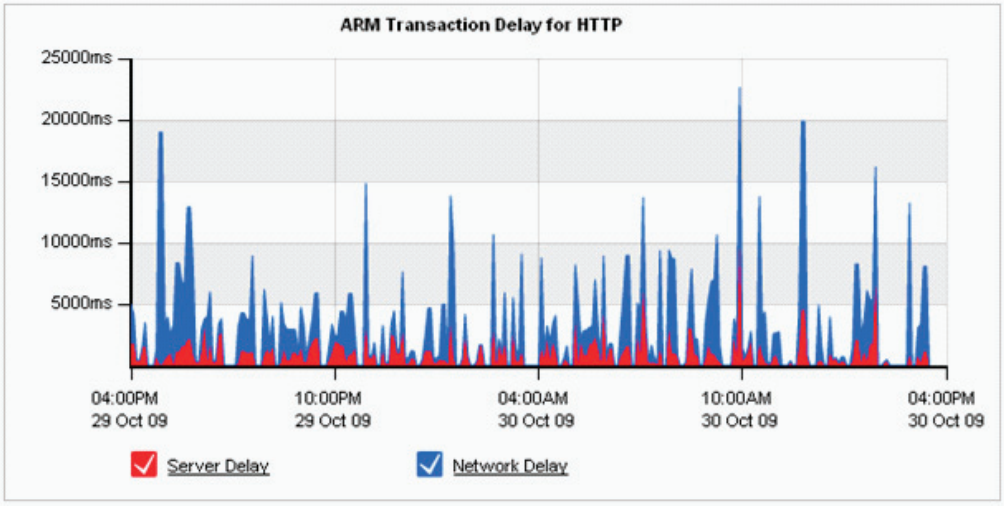


Figure 1: Application response trending data.

When the aforementioned network visibility technologies are implemented, they result in benefits such as:

- Paint a clear picture of overall bandwidth usage by traffic type and user
- Identify traffic including evasive applications like P2P
- Trend utilization to the application and user level
- Monitor performance of critical applications and servers and gain actionable information
- Determine behavior and align it to mission critical needs
- Develop a foundation for making well informed decision on how to best improve application performance, reduce costs, and eliminate waste

Policy Based Traffic Shaping

Policies provide bandwidth guarantees to critical applications to ensure they receive the network resources they need in order to operate effectively. Policies also, limit and constrain recreational or non-business critical applications. This minimizes the negative impact non-sanctioned applications can have on the network. In addition, traffic discard or blocking can allow the IT organization to completely block applications if required. This is very effective in reducing congestion and protecting an organization from the dangers of file sharing, P2P, and other rogue applications that may appear on a network.

Policy based shaping helps prevent network congestion by ensuring network services align with the policies set by the network manager to meet user expectations. To deliver the application performance guarantees in today's networks several different technologies must be employed together. When control is decentralized, computers and the applications they run will attempt to achieve the maximum available bandwidth. The result is many non-critical applications consume resources at the expense of the applications an organization uses to do business. In order to shape traffic several capabilities are required:

- Sort traffic into policies using deep packet and heuristics traffic inspection
- Apply performance controls to each policy - bandwidth guarantee and max limits
- Schedule packets in order to fairly allocate resources to policies
- Order packets on the wire to achieve performance guarantees – prioritization

When policy shaping is employed organizations are able to improve the utilization of existing bandwidth. Controlling the non critical applications and enforcing bandwidth guarantees will increase application performance and provide faster end user response times to critical applications. This ultimately results in increased user productivity and reduced help desk calls. It will also aid in eliminating the undesirable traffic form the network maximizing the performance of the WAN.

Where does an organization begin?

The first step an organization should take when they begin evaluating traffic shaping is to assess their organizational needs. Are there specific business initiatives such as IT cost containment or reduction, controlling losses in productivity, or prevention of poor user experiences the organization is trying to achieve. Does the network need to support a new application rollout, or new staff members, or additional branch offices.

Below are several business initiatives organizations achieve when implementing traffic shaping:

Cost Containment	Support New Initiatives	Network Visibility & Reporting

• Control Losses in Productivity
• Mission Critical Applications impacted by non-critical traffic
• Slow Response Times
• Prevent Poor User and Customer Experiences
• Allocate Bandwidth Intelligently
• Manage Overall Network Cost of Ownership
	• Next Generation Networks
	• Technology Convergence
	• New Application Rollout
	  • VoIP
	  • Citrix
	  • eLearning
	  • Collaboration
	  • Video Conferencing
	  • Disaster Recovery
		• Visibility into Application Performance
		• Real-time diagnostics
		• SLA Reporting
		• Trending and Analysis
		• Proactive Approach to Network Management

Next, an organization should analyze their current wide area network architecture. How much bandwidth does each WAN link support? What is the cost of each of these links on a monthly and annual basis? Are the WAN links currently at full utilization?

Finally an organization should understand the applications that are critical to the organizations business. Is voice over IP or video conferencing being used for communications? CRM for customer relationship, file transfers, email, and database traffic. Do applications have slow response times? Are users complaining about applications or the network? Are these critical applications being impacted today?

Categorize Your Traffic

Once an organization has implemented visibility into the behavior of their network, applications, and bandwidth usage, it is time to categorize the traffic. Most organizations have applications that fall within the following traffic categories:

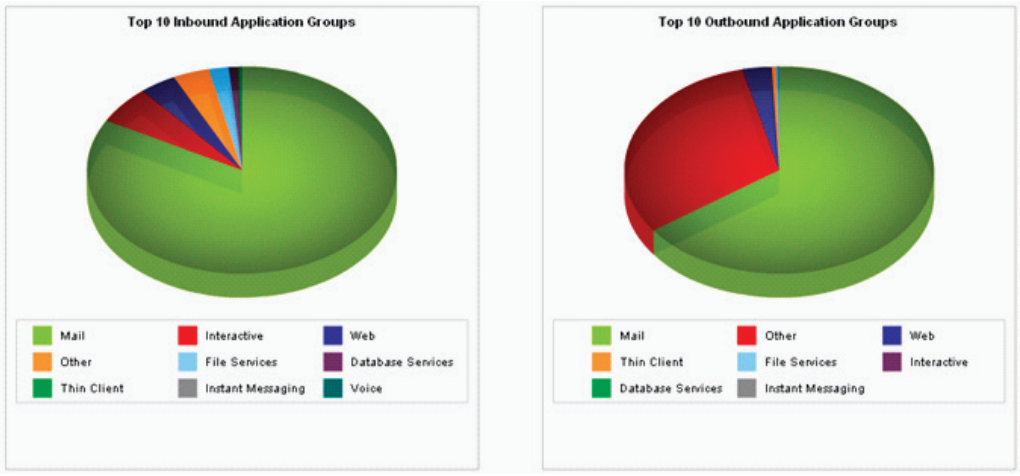


Figure 2: Inbound and outbound network traffic categories.

	Real Time Communications	Business Critical	Bulk Data	Undesirable
Expectation	Traffic is expected to be delivered on-time. Delay are unacceptable and performance can easily be negatively impacted.	These are the bulk of an organizations key applications. They require appropriate network resource to operate efficiently.	Can consume large amounts of network resources at the expense of sensitive traffic. Usually not time sensitive but does need guaranteed bandwidth and performance improvements.	Typically consumes large amounts of network resources at the expense of all other traffic. Usually not business related. A typical example would be peer 2 peer or recreational applications.
Application	• Voice Over IP • Video Conferencing • Distance Learning • Collaboration	• Thin Client • eLearning • CRM/ERP • Database • Interactive real-time transactions	• FTP • Email • Files • Software Updates	• BitTorrent • Streaming Media • Gaming • Internet Radio
Requirement	Need precise bandwidth guarantees to operate without problems.	Need bandwidth guarantees to operate without problems.	Need control and optimization techniques to minimize their network footprint and also speed their performance.	Need highly accurate visibility and ability to limit or squeeze off network.

Establishing Baseline Performance

Establishing a baseline of acceptable network performance and behavior is critical. It will provide the network administrator a clear picture of the starting point which will allow them to compare the impact of traffic shaping and help serve as a justification for expenditures. The baseline should include an understanding of the available network bandwidth, what applications are running on the network, the typical behavior of these applications and their response times, and network and server based latency. Proper baselining allows network administrators to make conclusions about application performance improvement, easily identify behaviors that are outside the acceptable range for network performance, and to differentiate between normal and anomalous traffic patterns and activity on the network.

Implementing Control Policies

The next step is to implement the appropriate traffic shaping and optimization policies to improve network and application performance. An organization should begin with a simple set of policies that address the various needs of the categorized traffic on their network. Once an initial set of policies is implemented an organization can begin to determine the effectiveness of the policies by monitoring and collecting information about the application behavior. Once this assessment is made each policy adjustment should be implemented and evaluated for its effect prior to making subsequent policy adjustments.

Figure 3: An example of traffic shaping policies.

Continuous performance monitoring is key to ensuring the traffic shaping is having the desired effect. Measuring response times for key applications after the implementation and comparing to the baseline information collected will allow an organization to evaluate the impact of the initial policies. It is also critical to view how traffic is passing through policies to determine if certain traffic types require additional bandwidth reserved or if they require less than already allocated. User feedback is also plays a critical role in determine the impact of policies. Power users should be questioned regarding the performance of applications and the use of the network.

Summing Up

Server consolidation and virtualization, as well as data warehousing and migration strategies has significantly increased network traffic. Organizations need an unprecedented level of visibility and control to manage it effectively. At the same time, there has been a surge in the number of mobile workers with smart phones, further increasing internet based application usage. Along with the legitimate increase in network usage, many organizations are grappling with how to effectively manage non-business traffic being generated by peer-to-peer and recreational applications.

Robust technologies exist that allow an organization to fully realize the potential of their network resources and achieve significant cost savings, defer upgrades, while improving employee productivity and satisfaction.

About Exinda

Exinda is a global provider of WAN optimization and application acceleration products. Exinda has helped over 2,000 organizations worldwide reduce network operating costs and ensure consistent application performance over the WAN. The Exinda Unified Performance Management (UPM) solution encompasses application visibility, control, optimization and intelligent acceleration – all within a single network appliance that is affordable and easy to manage.

Founded in 2002, Exinda is headquartered in Boston, Massachusetts with regional offices in Canada and the United Kingdom. Research and Development is centralized in Melbourne, Australia.

To learn more about Exinda's award-winning solutions, contact your local reseller or visit www.exinda.com.

exinda

North America
+1 877 439 4632

EMEA
+44 808 120 1996

Asia Pacific
+61 3 9415 8332

www.exinda.com