Traffic Management Overview

CONTENTS

Overview

Today, the network (Internet, Intranets and Extranets) is playing an ever increasing role in business and in applications. Companies like Amazon.com and Yahoo depend on Network access to conduct 100% of their business. The applications that run such leading edge companies require that bandwidth be available. Traffic Management, QoS, allow an organization to manage bandwidth access which is one of the most costly resources for an origination and one that requires a cost effective way to manage.

Advances in network technologies and applications introduce significant traffic management challenges. The demand for higher network performance has forged a new business requirement. Network access must be delivered in a reliable manner; otherwise the network cannot provide the infrastructure required delivering information expected by users.

Traditionally, the network tries to allocate as much bandwidth as needed until the bandwidth is exhausted. Mission-critical needs have to compete with employee web browsing. The Network becomes crowd, slow in responding and packets get dropped. The result translates to lost business opportunities and lost revenues.

We begin with a quick review of the business needs for QoS. We then describe the technology solution and examine how NetScreen has implemented this in a policy based security appliance, integrating QoS with firewall and VPN, Virtual Private Networks.

Business Need

The challenge in today's networks is to provide consistent and reliable performance levels that are acceptable in today's business applications. The traffic patterns change consistently at a rapid pace, and are unpredictable. One example is audio and video streaming, without adequate traffic management, there won't be enough bandwidth to support these media rich content. Another example is push applications, like Pointcast, which transfers large bursts of packets across the network, exhausting a large portion of the limited network resources. Multimedia applications generate heavy traffic, and usually generate large amounts of data due to a download request. The solution to this is an effective traffic management solution that can assign priorities to specific applications, so that the more important users and/or critical services are not starved of network bandwidth.

Many corporations today are utilizing Virtual Private Networks (VPN) to communicate between offices in a Wide Area Network (WAN). A VPN secures the transmission of data across the public network through the use of encryption. Bandwidth management must be able to effectively handle the large volume of traffic carried by VPNs. Otherwise, the VPN will not be used as it was initially intended. VPNs should have as little impact as possible over the network. Traffic shaping ensures VPN traffic does not slow down overall network performance.

Technology Solutions

Traffic shaping allows network managers to "shape" traffic according to policies based on connection speed, and IP address. Based on these parameters, traffic shapers control how IP traffic moves from the LAN to the WAN.

Proper traffic management ensures performance of mission critical applications without sacrificing lower priority applications. It is designed to eliminate burst and delay effects inherent to Internet traffic. Network traffic is tuned using queuing and supplying maximum rate of transmissions. This approach optimizes network performance and alleviates network congestion.

An ISP may need to provide limited resources that are shared by multiple clients. They may have specific clients who pay for a specific bandwidth and the ISP must ensure that the customer has the bandwidth according to the rate promised. This business model dictates higher priority for contractual obligations, rather than hardware requirements.

"If you need to process at line speeds of T3 [45 Mbits/s] or higher, you'd need the performance of a custom hardware chip", says Melinda LeBaron, research director for Dataquest/Gartner group. Netscreen agrees with this assessment, and includes a hybrid design that combines the performance of a custom hardware chip with the flexibility of a software implementation. The benefit of the hybrid design is the flexibility to upgrade software while maintaining the robust speeds required for today's demanding network hardware.

Technology Review

Traffic Shaping typically operates in one of two schemes: queuing or TCP/IP rate control. Netscreen's implementation uses a combination of queuing and TCP/IP rate control.

Queuing works best for applications with bursty traffic like multimedia AVI files. Queuing categorizes traffic and sets policies based on priority ranking. The data is placed into a queue and emptied based on the order of importance. The lower priority queues can only be transmitted after the higher priority queues have been transmitted.

The drawback with queuing is that if policies aren't set correctly, low priority traffic will be left with little or no bandwidth, and the result could be dropped packets, and numerous re-transmissions. This can make congestion worse than with no traffic shaping.

The second method of traffic shaping makes use of TCP/IP rate control. Here, the TCP/IP application can signals the source to send fewer packets if the network is becoming congested. The source decreases its window size according to the request. TCP/IP rate control was designed to handle the different network bandwidths available.

Many applications use TCP to guarantee delivery, and they take advantage of TCP's windowing to burst data across the network. To throttle down the bandwidth, Netscreen re-writes the TCP windowing information as traffic goes through. Typically, TCP informs the source to increase or decrease the TCP window size, depending on the characteristics of the network. However, rate control continually adjusts the interval at which the source receives the windowing information. The size of the window is always under control (instead of waiting for congestion to dictate the window size). This regulates the rate at which the packets are sent.

Netscreen Implementation

Traffic shaping is available in the Netscreen 10 and Netscreen 100 in release 1.00 and above. Traffic shaping can be configured as an option for each policy defined in the firewall. Traffic and packets are controlled and shaped according to the configuration of each policy. Traffic shaping in the Netscreen is controlled by three parameters: maximum bandwidth, guaranteed bandwidth, and priority.

Maximum bandwidth is the combined maximum throughput of all session from a policy. Traffic beyond this bandwidth will be dropped. Guaranteed bandwidth is the combined guaranteed throughput. Traffic below this will be transmitted without any traffic shaping mechanism. Priority controls the handling of traffic that falls between the maximum rate and the guaranteed rate. Traffic with higher priority will be passed with higher probability. A total of 8 priorities are possible on the Netscreen.

The maximum bandwidth of each policy may not exceed the configured interface bandwidth. The total guaranteed bandwidth might not exceed the configured interface bandwidth. However, the maximum bandwidth of any policy may exceed the actual

interface bandwidth. In this case, the actual traffic will be limited by the actual interface bandwidth.

The total guaranteed bandwidth might exceed the actual interface bandwidth. All guaranteed bandwidth will be prorated according to the actual interface bandwidth. The total guaranteed bandwidth could also be less than the actual interface bandwidth. In this case, excessive interface bandwidth will be given to traffic that needs extra bandwidth, with precedence given to higher priority traffic.

Conclusion

NetScreen's traffic shaping engine is completely integrated into the firewall. Security policies can define the traffic management schemes. Netscreen furnishes bi-directional traffic shaping, which performs traffic shaping from LAN to WAN, as well as WAN to LAN.

NetScreen's implementation allows an administrator to proactively manage bandwidth allocation. You can define and deliver specific levels of quality of service for nodes accessing an internal web server, and internal users accessing external resources like the Internet. By categorizing traffic and assigning bandwidth management, you can deliver packets with smooth, predictable service.