

Optimizing Performance for Voice over IP and UDP Traffic

OPTIMIZING PERFORMANCE FOR VOICE OVER IP AND UDP TRAFFIC

Voice-over-IP (VoIP) is a special category of network traffic that requires separate and distinct consideration from the more traditional TCP-based traffic found in IP networks. Certainly, user expectations about the quality and availability of telephony services are at a much higher level than for web and file applications. Due to these elevated user expectations, VoIP traffic must be handled with care. Improper configuration and support of VoIP traffic can lead to degraded voice quality and dropped calls.

Some WAN optimization vendors have been making noteworthy claims regarding their ability to “accelerate” and “improve” VoIP traffic. The vendors suggest that they have exclusively discovered the technology to deliver significant compression results for VoIP traffic *while also* improving voice quality. But just how credible and reliable are these claims and statements? While they certainly sound interesting, these claims are simply not supported by the facts.

VoIP has Special Challenges

VoIP is considered real-time traffic; VoIP datagrams must be delivered by the network with a minimal amount of jitter in order to maintain voice quality. Latency must be kept below 150ms if possible. Any deviations from these standards will affect voice quality. For these reasons, VoIP traffic must receive premium network service through any IP network, and any techniques employed to expedite delivery of VoIP traffic must not add latency or jitter.

To “accelerate” VoIP traffic is an ambiguous, often-misleading phrase. VoIP traffic moves at the speed of light, just like all other types of IP traffic. It can be delayed by network congestion and various other issues, resulting in degraded voice quality and dropped calls. But to “accelerate” VoIP traffic does not make sense in the context of the network (you cannot make VoIP packets travel faster than the speed of light) or the telephony application (you cannot make the participants talk any faster).

The remainder of this document will explore the various means used to expedite and optimize delivery of VoIP traffic across WANs. Various product vendors offer each of these techniques – some approaches are worthwhile, while others are of questionable value at best.

Compressing VoIP Traffic

Each VoIP conversation does not represent a significant amount of traffic – a few kbps or so. While many vendors may draw attention to their “impressive” compression results for VoIP traffic, it is important to remember that in most WANs, the baseline byte-volume of VoIP traffic pales in comparison to the huge volume of other TCP/IP traffic. Even if significant compression results could be achieved, there will not be a significant amount of bandwidth saved when considering the total volume of all traffic in the IP network.

VoIP uses IP datagrams to transport analog data that has been encoded into binary format through efficient algorithms. These algorithms have been developed after many years of engineering effort, and the resulting encoded data generated from a normal phone conversation has little or no repetition and entropy. As a result, there are very few opportunities to compress the VoIP datagram payload without affecting voice quality.

Instead, the primary opportunities to compress VoIP traffic come from a technique known as header compression. Header compression attempts to leverage the repetitive byte values that are found in the VoIP packet header. One distinguishing characteristic of VoIP that makes this somewhat worthwhile is the significant size of the header relative to the size of the entire datagram, especially when compared to “normal” TCP/IP traffic. However, the opportunity is limited because the size of the header is rarely larger than the size of the VoIP payload, and claims to be able to deliver greater than 30% compression through header compression techniques simply do not add up.

In addition, header compression technology is integrated into most WAN routers. If the limited gains from header compression are worth pursuing, then a separate WAN optimizer product is not necessary.

Furthermore, header compression techniques that are based on tunneling and coalescing of packets will add jitter and latency to the end-to-end VoIP traffic. That is because some packets are delayed so that they can be buffered up and coalesced with other packets through use of common tunneling headers. While these techniques can effectively reduce and eliminate some VoIP headers, the added jitter and delay will affect the quality of the voice conversation. The greater the aggressiveness with which these compression techniques are applied, the greater the resulting jitter and delay, and negative impact on overall voice quality. As a result, many network administrators simply choose not to apply header compression techniques due to their impact on the voice application.

Quality of Service (QoS)

QoS policy enforcement is an important mechanism to expedite delivery of VoIP through shared IP networks. It should be deployed on the WAN router if possible, because the WAN router is the bottleneck that must match the high-speed LAN to the low-speed WAN connection. Consequently, the WAN router is the point at which network congestion occurs, and where packets get dropped or delayed. This allows the WAN router to prioritize VoIP over data-oriented TCP/IP packets and prevent jitter and loss for voice packets when network congestion occurs.

Some vendors market QoS policy enforcement capabilities within their own WAN optimization products, allowing QoS configurations to be offloaded from the WAN router. The idea is to have the WAN optimizer perform the traffic shaping and packet dropping so that the WAN router never experiences network congestion, and does not drop packets. In a limited number of cases there may be reasons to perform QoS policy enforcement in this manner. Some of the advantages of this approach include:

- Superior graphical interface compared to the CLI found in the Cisco WAN router
- Concern about hardware-related processing constraints in the WAN router
- More granular control from the ability to do layer 7 application-level inspection

Furthermore, in some situations, the customer may not control or have access to the WAN router which is managed by a service provider. In these instances, performing QoS policy enforcement on a separate device such as the WAN optimizer becomes the only alternative.

However, in a large number of cases it will not be possible for the WAN optimizer to prevent network congestion. Specifically, this is true when the WAN optimizer does not observe and control all traffic entering the WAN router. When this is the case, there will be consequences for performing QoS policy enforcement exclusively in the WAN optimization device, because it cannot know the state of the WAN router's packet buffers. If the WAN optimizer has no direct knowledge of whether or not network congestion is actually taking place, any QoS decisions on whether to drop, delay, or expedite a given packet will be made with incomplete information.

This concept is illustrated in Figure 1 below. In the most simple trivial network where there is one LAN connection and one WAN link, the router's only role is to match the slower-speed WAN to the high-speed LAN. In this case QoS can be configured on the WAN optimizer to observe and shape all traffic entering the bandwidth-constrained WAN.

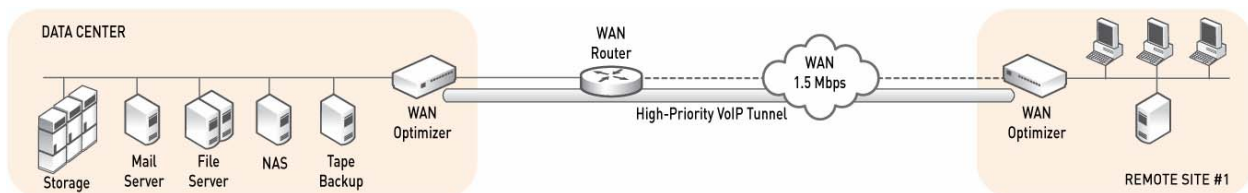


Figure 1 – In the most trivial network with only one WAN link and one LAN connection, the WAN optimizer can effectively control and shape traffic so that the WAN router never sees network congestion.

On the other hand, in a more complex network with multiple WAN links, it becomes more difficult for the WAN optimizer to observe and regulate traffic seen by the WAN router. The diagram in Figure 2 below illustrates traffic sent from Remote Site #1 to Remote Site #2, which traverses the WAN router without passing through the WAN optimizer. Because the WAN optimizer cannot “see” this traffic, it cannot control nor account for it when executing QoS enforcement policies.

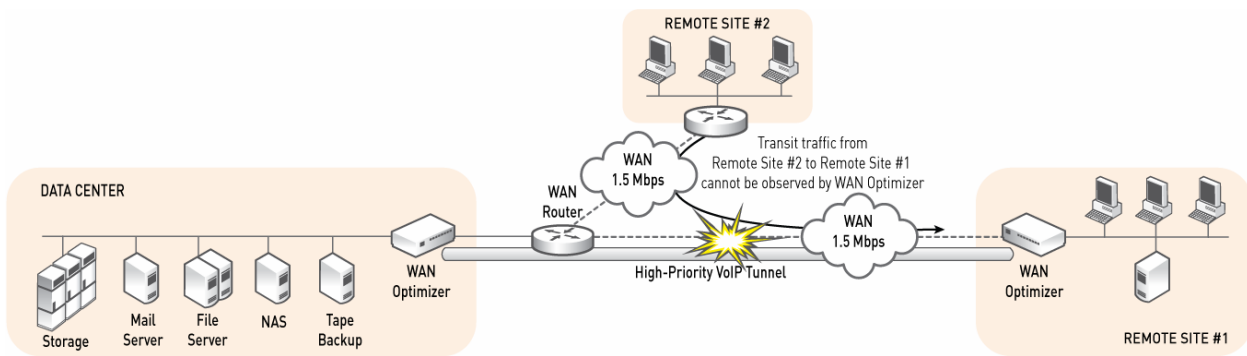


Figure 2 – In a more complex network with more than one connection link supported by the same WAN router, the WAN optimizer cannot observe all traffic. If the WAN router is not configured with QoS policies, then high-priority traffic could encounter network congestion, resulting in delay and loss of packets classified as high-priority by the WAN optimizer.

More seriously, vendors that aggressively market the QoS enforcement capabilities of their WAN optimization products often fail to communicate to their customers the reasons for configuring QoS on their WAN routers. As a result, some customers may be unaware that their mission-critical VoIP traffic is vulnerable to loss and delay.

Adding WAN Bandwidth through your Telecom Provider

VoIP traffic consumes a relatively small and predictable amount of bandwidth. Adding WAN bandwidth should be considered if VoIP call quality and reliability are an issue or concern. Nothing comes for free, and the appropriate investments must be made to create a suitable network infrastructure to support your voice traffic. While this approach may increase telecom infrastructure costs, the amount of bandwidth to support a given call is not that significant in the first place – a few kbps per call. Since VoIP is not bursty, one can easily predict and plan for the amount of WAN bandwidth needed to support a given number of telephony conversations.

Adding WAN Bandwidth through Riverbed

TCP/IP traffic sharing the WAN link with VoIP represents a significant amount of raw traffic – often many times the amount of bandwidth consumed by the VoIP traffic. Unlike VoIP, this traffic contains significant amounts of repetitive information that is transferred over the WAN numerous times by various TCP/IP-based applications. Riverbed's Scalable Data Referencing (SDR) technology has demonstrated the ability to consistently reduce the volume of this traffic by between 70% and 95%. As a result, significant amounts of WAN bandwidth resources are made available for voice and other UDP-based applications.

Use of Riverbed Steelhead appliances to free up bandwidth for VoIP traffic has the following advantages:

- **Avoids increased telecom costs** – Riverbed Steelhead appliances involve a one-time non-recurring investment.
- **Uses existing network infrastructure** – No need to upgrade WAN routers or purchase new line cards.
- **Integrates with QoS** – Riverbed Steelhead appliances are transparent to QoS policy enforcement in the WAN routers through DiffServ.
- **Accelerates TCP traffic** – Riverbed Steelhead appliances not only helps VoIP traffic by freeing bandwidth resources, but CIFS, MAPI, and most other TCP-based traffic will experience increased response-time performance.

As previously discussed, compressing VoIP traffic is usually ineffective and leads to unwanted consequences. Consequently, adding WAN bandwidth – either through your telecom provider and/or through use of Riverbed Steelhead appliances – is often the only effective recourse for supporting a growing number of telephony conversations without negatively affecting voice quality and reliability.

Optimizing Delivery of Other UDP Traffic

Of the applications that transport data using UDP, only a select few use it over the WAN. UDP does not have embedded mechanisms to retransmit lost or reordered packets, and UDP-based applications that transfer large amounts of data (such as TFTP and NFSv2) are unreliable when used over the WAN. Consequently, very few people even attempt to use these applications over a WAN infrastructure.

Rather, UDP traffic that is commonly seen over the WAN is the signaling and messaging-type application such as SNMP, DNS, and IP multicast. This traffic does not represent a significant amount of data, especially when compared to the overwhelming amount of TCP traffic found on most WANs. Compressing this traffic therefore yields negligible overall benefits.

However, these applications can be affected by packet loss that results from network congestion encountered on the WAN. This can affect the responsiveness of the application, which is then forced to re-send the lost data. Fortunately, it is possible to address the causes of network congestion to improve the reliability of UDP, in the same manner as we discussed with VoIP traffic.

Use of Riverbed Steelhead appliances can address this issue by eliminating 70% to 95% of the TCP traffic that otherwise would cause network congestion. This change can make delivery of UDP datagrams more reliable, and thus allowing the application to perform more consistently without having to re-issue lost UDP datagrams over the WAN.

Summary

Voice-over-IP traffic requires special considerations that are separate and distinct from traditional optimization approaches applied to TCP traffic. VoIP is typically non-compressible data, and header compression and packet coalescing approaches almost always lead to degradation in voice quality. Alternatively, QoS enforcement approaches are very effective for prioritizing VoIP traffic over WANs, particularly when QoS is enforced in the WAN router.

However, the most basic requirement for VoIP traffic – bandwidth – must be addressed as well. If adequate bandwidth does not exist, then VoIP quality will suffer. Bandwidth can be addressed by adding more bandwidth through the telecom provider, or through use of Riverbed Steelhead appliances. Using Riverbed Steelhead appliances, it is possible to avoid disruptive upgrades of WAN bandwidth and higher recurring costs paid to the telecom provider.

Riverbed Technology, Inc.
501 Second Street, Suite 410
San Francisco, CA 94107
Tel: (415) 247-8800
www.riverbed.com

Riverbed Technology Ltd. UK
200 Brook Drive, Green Park
Reading RG2 6UB
United Kingdom
Tel: +44 118 949 7002

Riverbed Technology Pte. Ltd.
350 Orchard Road #21-01/03
Shaw House
Singapore 238868
Tel: +65 68328082

Riverbed Technology K.K.
Shibuya Mark City W-22F 1-12-1
Dogenzaka, Shibuya-ku Tokyo
Japan 150-0043
Tel: +81 3 4360 5357

© 2006 Riverbed Technology, Inc. All rights reserved. Riverbed Technology, Riverbed, Steelhead and the Riverbed logo are trademarks or registered trademarks of Riverbed Technology, Inc. Portions of Riverbed's products are protected under Riverbed patents, as well as patents pending. WP-VOIP033106