Managing SIP-based Applications With WAN Optimization

Worry-Proof Internet
2800 Campus Drive · Suite 140 · Plymouth, MN 55441
Phone (763) 694-9949 · Toll Free (800) 669-6242

Managing SIP-based Applications With WAN Optimization

Building Redundancy and Failover into SIP-Enabled Networks

Communications has always been the foundation for commerce, and today businesses have a variety of communication options. The most powerful form of communications is still face-to-face meetings. However, it is not realistic in terms of cost and logistics, except for the most important interactions. Electronic communication via email and instant messaging are the most convenient and cost-effective, but they lack intimacy.

Yet, despite all the technological advances in communications, the greatest balance of efficiency and intimacy is still the telephone, and why businesses place so much importance on their voice networks. However, there are many issues that make managing voice network reliability more difficult than ever.

Central to this is the fact that today's IP-based networks are more complex. In addition to legacy POTS, companies now have VoIP and wireless voice networks. Straightforward voicemail has evolved into complicated integrated messaging platforms, and the traditional handset has been replaced by devices such as soft (IP) phones, cellular phones and smart phones.

Additionally, when it comes to voice networks, the demands for availability are higher than ever before. Businesses require the network to work without interruptions. The cost of losing phone service has such a devastating effect on business it requires five nines of reliability.

Today, one of the most critical issues businesses face is how to effectively manage their voice networks, and the potential for integrating their voice systems with multimedia and new interactive capabilities. It is within this challenge that there lies an opportunity for service providers, system integrators, VARs equipped with the appropriate solutions to create new revenue-generating services.

The Growing Deployment of SIP Communications

The age of disparate voice and data networks is slowly, but surely coming to an end. Single, converged networks are beginning to deliver all forms of communications. In addition to reducing costs, this convergence can enable a great diversity of new services.

At the center of this transformation is Session Initiation Protocol (SIP), an IP telephony standard developed to manage the setup and tear-down of voice-related IP networking sessions. As voice, data, and video communications become more unified, IP-based applications will increasingly use SIP to manage the connections.

SIP enables communication devices such as fixed and mobile phones to interoperate with Internet services such as email, the web, instant messaging, and multimedia collaboration. Beyond VoIP, SIP-based applications can include many innovative services, such as video conferencing, streaming media distribution, online gaming, voice-enriched e-commerce, and other services.

With the increasing deployment of SIP technology, Telcos and service providers need to deliver five nines of service availability and quality application delivery. Similarly, the need for carrier-grade service requirements has placed the same demands on equipment providers and the VARs and system integrators charged with deploying these solutions.

When it comes to delivering the high-availability and performance needed for applications that use SIP, fundamental problems need to be addressed that require network redundancy and failover to ensure calls remain connected, even in the case of a service provider outage. To be specific, problems with NATing within a SIP-enabled network must be addressed to ensure true redundancy for SIP traffic going over multiple WAN links.

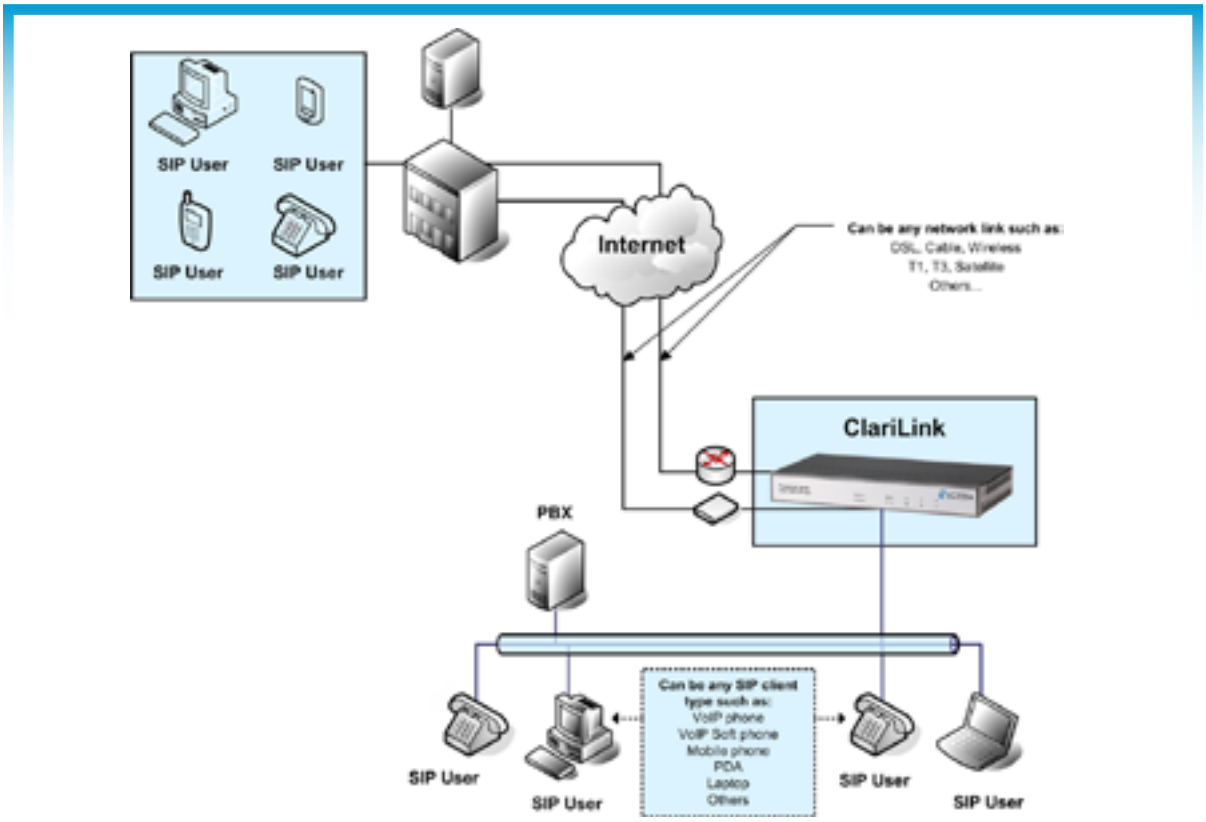
SIP and Network Address Translation (NAT)

There are several areas within a SIP-enabled network where NAT and Remote Transport Protocol (RTP) become

a problem. SIP packets go from users (with NAT) from their private (un-routable) IP addresses coded into the message headers. However, the SIP packets may not get processed by a NAT device over the WAN. NATing operates only on IP packets. Therefore, when the SIP packets arrive at their destination, they are processed and responded to with unusable source address information. Responses to requests cannot be routed back to the originating party, as the addressing information is not globally routable.

Additionally, Session Description Protocol (SDP) messages used to negotiate the session format (codecs, ports, IP's, etc.), are often enclosed within the body of the SIP message. However, SDP messages are not processed by a SIP proxy (according to IETF standards). Therefore, they will contain non-routable contact information.

Ecessa's VoIP Failover and Monitoring service has addressed the issues having to do with the NATing of SIP traffic –enabling SIP traffic to go over NAT environment successfully. The ability to provide failover and load balance SIP connections over multiple WAN links is critically important for successful, scalable VoIP environments. Ecessa is at the forefront of delivering this essential capability, and unique features such as integrated NAT, SIP Proxy and SIP registrar reinforce Ecessa's position as a true value/performance leader in supporting the unique and challenging requirements of enterprises of all sizes, Telcos, Internet service providers, and system integrators.



This diagram shows ClariLink managing SIP traffic for both LAN and WAN traffic

The Need for VoIP Failover and Monitoring Services

VoIP Failover and Monitoring Services enabled on a micro-appliance allow it to function as a NAT device and SIP Proxy within the network, providing a standard approach to ensure reliable SIP service availability across multiple and diverse WAN connections. While non-SIP-enabled WAN link solutions may be helpful in solving general link load balancing, there are inherent problems associated with failure in supporting SIP message flow, and interoperability issues associated with NATing SIP traffic.

VoIP Failover and Monitoring Services are required to perform SIP operations such as session management and routing, transport conversion and security functions. By acting as a NAT device and SIP Proxy the micro-appliance

can perform these operations, and avoid NATing interoperability issues. A micro-appliance with VoIP Failover and Monitoring enabled functions as a default gateway for the traffic to be routed through it. This enables the routing of all traffic through the micro-appliance regardless of SIP signaling, media or other types of traffic. Such a micro-appliance can be configured as an inbound Proxy for all traffic appointed by the DNS servers, or peering SIP elements, as well as the outbound Proxy for all traffic going out from the LAN.

VoIP Failover and Monitoring Services Deliver WAN Link Redundancy for SIP Deployments

In much the same way as dedicated global load balancers provided high-availability, scalability and performance for web applications in the late 1990's, VoIP Failover and Monitoring Services are delivering the same ability to scale, deliver high-availability and reliable communications for VoIP and other applications that are SIP-enabled.

For enterprises, Telcos, service providers, VARs and system integrators, such services can dramatically simplify and reduce SIP application deployment time. VoIP Failover and Monitoring Services should provide the following capabilities:

- Deliver seamless call failover for inbound and outbound connectivity
- Enable SIP devices to work over multiple, diverse WAN connections
- Eliminate problems associated with NAT
- Provide link load balancing for SIP traffic

What to look for in VoIP Failover and Monitoring Services

The list below is the criteria that a small-to-medium sized enterprise should look for when choosing a solution for their needs.

High Availability (Hot Standby) - Since all outbound and inbound traffic must pass through the micro-appliance, should it fail, the entire site will be inaccessible. To address this, some vendors support redundant configurations. Usually, a standby (or redundant) configuration is supported - sometimes referred to as HA (High-Availability). Most sites utilize at least one HA pair - as it would be risky to deploy multiple WAN links for redundancy and scalability, only to lose the entire site due to the micro-appliance hardware failure. Should one of the micro-appliances go down, the monitoring service will notify the administrator.

Outbound Load Balancing and Failover – the WAN Optimization Services should provide outbound bandwidth load balancing and failover. The user defines weights (bandwidth capacity) based on the bandwidth of each WAN link. When a session or call is generated from the LAN, the appliance computes which link has the most available bandwidth and routes traffic from that session or call over that particular WAN link. The device typically allows the selection of two link load balancing algorithms:

Symmetrical round robin - routes sessions to all links in a round robin manner.

Intelligent (weighted) load balancing - computes a ratio between the weight (bandwidth capacity) of the different WAN links, and then routes sessions and calls accordingly. That is, the faster the link, the more sessions or calls that will be sent over that link, in order to make the most efficient use of all the bandwidth available. Additionally, an intelligent link load balancing solution will examine the amount of real-time traffic on each link, compared to the amount of available bandwidth resources left, and choose the best path for the next session or call's most optimal route for performance.

Inbound Load Balancing and Failover - is accomplished by the micro-appliance acting as the authoritative DNS server for the domain. The appliance advertises all available WAN links to the DNS caching servers which in turn resolve the domain names to queries in a round robin format. In this manner, all externally initiated sessions are load balanced over all available links. Since the appliance is resident at the domain site and is able to directly monitor the link status, failed links are removed from the DNS tables immediately upon failure. By setting the host name record Time-to-Live (TTL) to a short period (i.e. 30 seconds), the DNS caching servers will flush their address tables and will update them from the appliance regularly, and thus be informed when a link fails.

Cost-effective - A quality WAN Optimization Service should deliver easy and affordable WAN and/or ISP link aggregation, inbound and outbound load-balancing, failover, traffic shaping and application prioritization. You

may use two, three or however many WAN links and ISPs you need. This allows you to leverage low-cost links, eliminate link congestion and bottlenecks, and use the traffic shaping services to ensure minimum bandwidth to SIP-enabled applications.

You can take advantage of the cost of a consumer ADSL link, and get business connectivity at that price. Not only can you get the flexible capacity – you can also buy cost-effective links from multiple ISPs, so that if one link goes down, you can automatically switch over to the other links.

Through bundling (aggregating) multiple, diverse Internet links from one or more ISPs, a WAN Optimization Service reduces the need to purchase multiple and expensive high-speed links. This enables you to increase bandwidth by using cost-effective links without compromising up-time. In addition to managing scalability and redundancy, the device cost-effectively utilizes all available WAN bandwidth through intelligent link load balancing, with features such as quality-of-service routing. WAN Optimization Services should provide controls for how bandwidth is used to support SIP-based applications and connectivity. This allows you to take advantage of the most cost-effective ISP rates, while ensuring appropriate levels of bandwidth are available.

WAN Optimization Services should allow you to choose the WAN link performance/cost ratio that best fits your needs; provides you with complete service provider independence; and eliminates the complexity of network protocols such as border gateway protocol (BGP). The service's inbound and outbound load balancing supports two or more Internet connections and provides SIP-based applications with access to the total available pool of bandwidth. Load Balancing routes Internet sessions from congested links, to links with more available bandwidth. It also provides automatic failover of Internet sessions from failed links to functional connections to eliminate a point-of-failure.

WAN and ISP failover - When the service detects a link failure it should automatically update the DNS record for your domain so that the server requests are sent to the IP address of your alternate server or server cluster. WAN Optimization Services should also provide for device failover through its active/passive failover capability. This eliminates the chance of the micro-appliance being the single-point-of-failure.

Disaster Recovery/Site Redundancy - Many businesses need to redirect Internet traffic to a disaster recovery site should a catastrophe disrupt a main site. WAN Optimization Services have, in effect, reduced the cost to ensure that site failover and fallback occur automatically, and reliably, making this functionality practical and affordable even for the smallest businesses.

Traffic Shaping and Application Prioritization – Traffic shaping includes the ability to prioritize network traffic to ensure that adequate bandwidth is always available to specific bandwidth-intensive applications, especially during periods of congestion. Rules determine bandwidth minimums and maximums for specific types of traffic and use load balancing and automatic failover to direct this traffic to links with sufficient bandwidth. WAN Optimization Services should provide traffic shaping and application prioritization support for traffic based on defined rules.

Performance - Performance of applications over the WAN directly affects response time. This includes not only total average transaction time, but assures that users located at performance-challenged sites (such as branch offices) still receive the acceptable level of performance. Performance is an important criterion for any piece of networking equipment, but it is critical WAN Optimization Services, because datacenters are central points of aggregation. As such, the micro-appliance should support extremely high volumes of traffic transmitted to and from sites. A simple definition of performance is how many bits-per-second the device can support. While this is extremely important, in the case of a micro-appliance, other key measures of performance include how many WAN links, how fast the failover occurs from one link to another, number of supported concurrent sessions, number of domain names, and number of hostnames within each domain.

Security - More advanced WAN Optimization Services include built-in firewalling and security features to provide added security and lower the cost of support, maintenance and overall infrastructure complexity through device consolidation.

WAN Virtualization - WAN Optimization Services can support WAN Virtualization to bond multiple network links into a single high-bandwidth channel to ensure high-availability for applications. If one link goes down or degrades in performance, traffic is automatically directed to the best working links without interruption. WAN Virtualization is a form of load balancing which allows for stateful failover of traffic to the best performing links to ensure critical applications avoid problems that occur when they are stopped on one link and restarted over

another link. WAN Virtualization ensures that critical applications avoid failures, and are never adversely affected, even after brief disruptions.

Ecessa's WAN Optimization as a Service (WaaS) delivers industry leading price/performance value

With VoIP Failover and Monitoring Services enabled, micro-appliances support a built-in SIP registrar and "session-aware" SIP proxy that sets up calls by rewriting private IP addresses and ports to public IP addresses and ports, for controlling the management of connections and IP addresses within a SIP-enabled network.

Providing reliability, high-performance, flexible scalability and high-availability to SIP-based networks, VoIP Failover and Monitoring Services perform both inbound and outbound link failover and load balancing of traffic among multiple SIP-enabled networks so that service availability is ensured even under high call volumes. SIP traffic is directed away from congested WAN links and service provider outages, providing increased reliability for SIP deployments. Firewalling and Security Services add protection and creates a single point of contact for outside SIP traffic, by hiding downstream proxies and SIP clients from the outside.

WAN Optimization as a Service (WaaS) is an affordable, yet feature-rich solution for SIP-enabled applications within a NATed environment. WaaS enables redundant WAN and service provider connectivity, and provides both outbound and inbound WAN and/or service provider load balancing and failover to ensure optimal WAN uptime.

WaaS' Traffic Shaping and Application Prioritation can be used to maintain multiple links separately and allocate Internet traffic across them. With the WAN Virtualization Service you can use bandwidth aggregation that combines multiple WAN links into what is effectively one large network connection.

VoIP Failover and Monitoring Services ...

- Ensure each SIP device user gets the best network experience possible over the WAN
- Enable optimization of SIP-enabled network traffic over the WAN according to throughput capacity
- Offer WAN Virtualization Services, providing uninterrupted Internet access for reliable performance for applications between sites
- Offer redundant hardware failover and monitoring capabilities for mission-critical applications to eliminate potential single points of failure
- Provide Traffic Shaping and Application Prioritization for bandwidth management that ensure SIP-enabled applications get the bandwidth required for optimal performance
- Allow for easy adding, removing, and managing multiple service provider connections
- Support any type of IP connectivity to meet increasing bandwidth needs
- Use multiple ISP and/or WAN links simultaneously, leveraging the total available bandwidth via link load-balancing, to maximize connectivity costs, while avoiding unnecessary link costs from underutilized back-up links

Summary

SIP deployments are continuing to increase, while due to their mission-critical nature, availability, scalability and security requirements have likewise become essential for their deployment. This advancement has resulted in the need for WAN Optimization Services. Using VoIP Failover and Monitoring Services on a micro-applicance that functions as a NAT device, SIP Proxy, and SIP registrar within the network will ensure SIP service availability, while providing an easy-to-deploy, manage and monitor solution for reliable service delivery. While server and global load balancers are appropriate for load balancing servers and dispersed sites, they do not provide the ability to create the automated failover and redundant WAN connectivity to support reliable SIP network connectivity.

Ecessa's WAN Optimization is the industry's first service that has solved the problems associated with NATing within a SIP-based network to deliver true link redundancy, failover and load balancing of SIP-based traffic over the WAN. Ecessa's WAN Optimization as a Service (WaaS) leverages 40 years of experience in delivering WAN connectivity solutions, to extend the same features and benefits to SIP service delivery, as with other applications delivered over multiple, diverse WAN links.