Application Note

How To Determine Bandwidth Requirements

08 July 2008

Table of Contents

Tested versions: Ingate Firewall/SIParator/MEDIAtor version 4.6.2

Revision History:

Revision	Date	Author	Comments
	2008-07-08	Scott Beer	1st draft

1 BANDWIDTH REQUIREMENTS

In a Teleworker, Home Office or Remote Office solution attention to bandwidth requirements is crucial to a successful deployment. Bandwidth requirements for the Voice and Video as well as the Data must me understood.

1.1 Voice Requirements

A Brief History Lesson:
Humans speak in Analog. This analog signal is transformed to digital through a process called PCM (Pulse Code Modulation). The PCM data flows in a continuous 'stream' of data bits at 64 kilo bits per second (kbps) or 64000 bits per second. Usual phone conversations occur in both directions simultaneously, thus 64kbps in each direction, 64kbps streaming audio up and another 64kps streaming audio down.

Voice over IP (VoIP):
Voice over IP (VoIP) moves this PCM voice information over an Internet Protocol (IP) network. This voice information is sent as data over an IP network in packets. In VoIP, the digitalized PCM voice is bundled into IP packets and sent out onto the IP network. Thus, the term Packet Switched telephony, now simply known as VOIP. Voice traffic now shares the same space as Data traffic.

Keep in mind that additional information needs to be added to the Voice Packet to travel over an IP network, such information as MAC address headers and IP/UDP/RTP header address information and more. This is about 54 bytes of information, depending on the network. (There are 8 bits per byte.) With VoIP, the transport medium has changed since the PCM voice traffic is now carried over an IP network, requiring a source, destination, User Datagram Protocol (UDP), and Real-time Protocol (RTP) headers. In the case of G.729, for example, 58 bytes are needed to transport 20 bytes of compressed Voice. With overhead like this, it is clear that TCP over Internet Protocol (TCP/IP) is not suited to transporting compressed voice (or very small packets). An alternative protocol is to use RTP header extensions in combination with UDP over IP. In addition, the PCM 'stream' needs to be 'packetized', so now there is a Sample Rate. Taking a continuous 64kbps Stream and sampling it down to 20 ms of 'sampled' audio. Sending now 50 packets per second to ensure the 64kbps of voice data. Keep in mind, the more frequent the sample rate the more network overhead is added, to the point where there will be significantly more overhead then data.

Compression Algorithms:
Standardized compression algorithms have since been developed for use in voice applications. Data Compression is the process of encoding information using fewer bits than an unencoded representation would use through use of specific encoding schemes. The most popular compression algorithms are G.723 (6 Kbps every 30 ms), G.729 (8 Kbps every 20 ms), and uncompressed G.711 (64 Kbps every 20 ms). The main differences between them are packet size, voice quality, and complexity.

1.1.1 Calculating VoIP Bandwidth

The bandwidth allocation is dependent on the codec or compression type used for the VoIP encoders. For VoIP using G.723 compression, a minimum bandwidth of at least 32 Kbps is required, while G.729 compression requires between 30-40 Kbps of bandwidth depending on the sampling rate (20 or 30 ms).

These calculations are:

- Voice packet size = transport header + IP/UDP/RTP header + voice payload size
- Voice packet size in bits = voice packet size * 8 [bits per byte]
- Voice packets per second = CODEC bit rate / voice payload size
- Bandwidth per call = voice packet size * voice packets per second

Here is a practical application of these formulas, assuming the use of the G.711 CODEC:

- Voice packet size (214 bytes) = MAC header (14 bytes) + IP/UDP/RTP header (40 bytes) + voice payload (160 bytes)
- Voice packet size in bits (1712) = voice packet size (214) * bits per byte (8)
- Voice packets per second (50) = CODEC bit rate (64 Kbps) / voice payload size (1280 bits)
- Bandwidth per call (85.6 Kbps) = voice packet size in bits (1712) * voice packets per second (50)

VoIP's impact on the network can be estimated by taking the bandwidth per call and extrapolating with the number of concurrent calls that are expected to be handled during peak busy-hour calling times, as previously calculated. Such estimates are usually made with the largest amount of bandwidth required by a CODEC, which is currently G.711.

Summary Throughput for VoIP Codecs

Codec (sample rate)	Single Direction Bandwidth
G.723 (30 msec)	~30 kbps
G.729 (30 msec)	~30 kbps
G.729 (20 msec)	~40 kbps
G.711 (20 msec)	~90 kbps

Once the call flows and call volumes have been identified and the available network bandwidth has been determined, the next step is determining how much bandwidth will be required to support the expected VoIP traffic.

Keep in mind the network connectivity, whether Full Duplex, Half Duplex or Synchronous or Asynchronous. As this has bearing on how much actual throughput is available. As Half Duplex and Asynchronous network connectivity do not have the same Bandwidth in both directions.

Additional VoIP Bandwidth Requirements Using VPN
IPsec is one common method for encryption. IPsec (IP security) is a standard for securing Internet Protocol (IP) communications by encrypting and/or authenticating all IP packets. IPsec provides security at the network layer. It is made of a set of cryptographic protocols for securing packet flows and key exchange. Cryptographic protocols such as Encapsulating Security Payload (ESP), which provides authentication, data confidentiality and message integrity; and Authentication Header (AH), which provides authentication and message integrity, but does not offer confidentiality. These additional cryptographic protocols add approximately 10 per cent overhead to the voice traffic over a VPN tunnel.

Summary Throughput for VoIP Codecs over VPN

Codec (sample rate)	Single Direction Bandwidth
G.723 (30 msec)	~33 kbps
G.729 (30 msec)	~33 kbps
G.729 (20 msec)	~44 kbps
G.711 (20 msec)	~99 kbps

Keep in mind the network connectivity, whether Full Duplex, Half Duplex or Synchronous or Asynchronous. As this has bearing on how much actual throughput is available. As Half Duplex and Asynchronous network connectivity do not have the same Bandwidth in both directions.

1.2 Video Requirements

There are many video multimedia compression formats, too many to discuss the bandwidth requirements of each in this document. Video Bandwidth requirements are dependent on a dynamic assignment of Resolution with the selected Compression Algorithm and by Frame sizes which all combine to create a 'Black Art" of understanding of Video compression and bandwidth requirements.

Here is a list of the video compression formats.

ISO/IEC	MJPEG · Motion JPEG 2000 · MPEG-1 · MPEG-2 · MPEG-4 ASP · MPEG-4/AVC
ITU-T	H.120 · H.261 · H.262 · H.263 · H.264
Others	AMV · AVS · Bink · Dirac · Indeo · Pixlet · RealVideo · RTVideo · SheerVideo · Smacker · Snow · Theora · VC-1 · VP6 · VP7 · WMV

Streaming video quality is dependent in part upon the video encoding process and the amount of bandwidth required for it to be viewed properly. Encoding a video so that it will stream at a speed for a user who has limited or low bandwidth, requires a high degree of compression is applied to both the video and audio tracks. This process eliminates portions of the audio and video data, which is why most videos you see don't appear or sound as clear as you'd like.

A user who is connected to the Internet using a high speed cable modem or DSL connection, for example, can watch any streaming video clip which has been encoded for transmission at their connection speed or lower. However, a user who is connected to the Internet using a 56k modem who attempts to view a streaming video clip which has been specifically digitized and encoded for transmission to cable modem users, for example, will get very choppy video which plays for a second or two and then pauses for several seconds until more video data is transferred to their PC.

Here is a small sample of the bandwidth possibilities using H.264 video compression.

Level number	Max macroblocks per second	Max frame size (macroblocks)	Max video bit rate (VCL) for Baseline, Extended and Main Profiles	Max video bit rate (VCL) for High Profile	Max video bit rate (VCL) for High 10 Profile	Max video bit rate (VCL) for High 4:2:2 and High 4:4:4 Predictive Profiles	Examples for high resolution @ frame rate (max stored frames) in Level
1	1485	99	64 kbit/s	80 kbit/s	192 kbit/s	256 kbit/s	128x96@30.9 (8) 176x144@15.0 (4)
1b	1485	99	128 kbit/s	160 kbit/s	384 kbit/s	512 kbit/s	128x96@30.9 (8) 176x144@15.0 (4)
1.1	3000	396	192 kbit/s	240 kbit/s	576 kbit/s	768 kbit/s	176x144@30.3 (9) 320x240@10.0 (3) 352x288@7.5 (2)
1.2	6000	396	384 kbit/s	480 kbit/s	1152 kbit/s	1536 kbit/s	320x240@20.0 (7) 352x288@15.2 (6)
1.3	11880	396	768 kbit/s	960 kbit/s	2304 kbit/s	3072 kbit/s	320x240@36.0 (7) 352x288@30.0 (6)
2	11880	396	2 Mbit/s	2.5 Mbit/s	6 Mbit/s	8 Mbit/s	320x240@36.0 (7) 352x288@30.0 (6)
2.1	19800	792	4 Mbit/s	5 Mbit/s	12 Mbit/s	16 Mbit/s	352x480@30.0 (7) 352x576@25.0 (6)
2.2	20250	1620	4 Mbit/s	5 Mbit/s	12 Mbit/s	16 Mbit/s	352x480@30.7(10) 352x576@25.6 (7) 720x480@15.0 (6)

Level number	Max macroblocks per second	Max frame size (macroblocks)	Max video bit rate (VCL) for Baseline, Extended and Main Profiles	Max video bit rate (VCL) for High Profile	Max video bit rate (VCL) for High 10 Profile	Max video bit rate (VCL) for High 4:2:2 and High 4:4:4 Predictive Profiles	Examples for high resolution @ frame rate (max stored frames) in Level
							720x576@12.5 (5)
3	40500	1620	10 Mbit/s	12.5 Mbit/s	30 Mbit/s	40 Mbit/s	352x480@61.4 (12) 352x576@51.1 (10) 720x480@30.0 (6) 720x576@25.0 (5)
3.1	108000	3600	14 Mbit/s	17.5 Mbit/s	42 Mbit/s	56 Mbit/s	720x480@80.0 (13) 720x576@66.7 (11) 1280x720@30.0 (5)
3.2	216000	5120	20 Mbit/s	25 Mbit/s	60 Mbit/s	80 Mbit/s	1280x720@60.0 (5) 1280x1024@42.2 (4)
4	245760	8192	20 Mbit/s	25 Mbit/s	60 Mbit/s	80 Mbit/s	1280x720@68.3 (9) 1920x1088@30.1 (4) 2048x1024@30.0 (4)
4.1	245760	8192	50 Mbit/s	50 Mbit/s	150 Mbit/s	200 Mbit/s	1280x720@68.3 (9) 1920x1088@30.1 (4) 2048x1024@30.0 (4)
4.2	522240	8704	50 Mbit/s	50 Mbit/s	150 Mbit/s	200 Mbit/s	1920x1088@64.0 (4) 2048x1088@60.0 (4)
5	589824	22080	135 Mbit/s	168.75 Mbit/s	405 Mbit/s	540 Mbit/s	1920x1088@72.3 (13) 2048x1024@72.0 (13) 2048x1088@67.8 (12) 2560x1920@30.7 (5) 3680x1536/26.7 (5)
5.1	983040	36864	240 Mbit/s	300 Mbit/s	720 Mbit/s	960 Mbit/s	1920x1088@120.5 (16) 4096x2048@30.0 (5) 4096x2304@26.7 (5)

2 VoIP Performance

VoIP has three specific performance requirements that have to be met in order to provide a quality voice conversation.

Latency

Anyone who has tried a long distance conversation over a satellite link may have experienced how excessive latency impacts voice quality. These long delays make it difficult for callers to have a natural speech patter to their conversation as it is difficult to determine when the person at the other end has finished talking. How much latency is too much? A rule of thumb is that one-way latency should not exceed 150 milliseconds. 150 millisecond delays are noticeable but tolerable, but when latency exceeds 250 milliseconds it becomes too difficult to conduct a conversation. Latency is a non-issue on the PSTN, but delays on IP networks can easily cause latency to exceed 150 milliseconds due to data and voice congestion.

End-to-end latency is the sum of encoding/decoding latency and transmission latency. The encoding/decoding latency introduced is proportional to the level of compression provided by the codec. For example, G.711 performs no compression and adds negligible latency while G.729 codecs compress voice to 8 kbps but add a one-way delay of about 25 ms of latency. Next is transmission latency where more significant delays can occur when voice packets are transmitted across a network, including LAN and WANs. Keep in mind that Voice and Data are sharing the same network and it is possible to for the voice packets get "stuck" behind data packets being sent over various WAN link types and link speeds.

Even when voice packets are not blocked by data packets they are subject to their own serialization delay – the amount of time that is takes to clock the bits onto a serial link. Again, this delay is determined by packet size and link speed. Reductions in packet size result in less serialization delay and therefore, lower end-to-end latency. The adjustment in the sample rate will adjust the packet size, but the more packets sent onto the network increases the overhead added to each conversation.

To avoid Latency, ensure reliable network connectivity. Planning and deploying a network with no points of congestion and adequate bandwidth throughout the network. Networks deployed as 'Flat' as possible with enough bandwidth to ensure proper data and voice communications result in the best performance. Avoid chaining multiple Layer 2 switches together in a long serial chain, as the higher the up the chain traffic goes the more congested the final up link becomes. Deploy and number of Quality of Service techniques to ensure the best prioritization of Voice and Data. Calculate the bandwidth requirements of the voice and data at all common points of convergence.

Jitter

Another key performance metric is jitter. Jitter is the amount of variation in latency that is experienced over time. SIP phones have some ability to buffer incoming audio streams to compensate for jitter, but excessive jitter can disrupt conversations. A rule of thumb is that Jitter should not exceed 50 milliseconds. Again, the PSTN has virtually no latency and therefore no jitter, but enterprise IP networks are subject to jitter caused by congestion on LANs and WANs and by packet buffering in routers and other network devices.

To avoid Jitter, ensure reliable network connectivity. Planning and deploying a network with no points of congestion and adequate bandwidth throughout the network. Networks deployed as 'Flat' as possible with enough bandwidth to ensure proper data and voice communications result in the best performance. Avoid chaining multiple Layer 2 switches together in a long serial chain, as the higher the up the chain traffic goes the more congested the final up link becomes. Deploy and number of Quality of Service techniques to ensure the best prioritization of Voice and Data. Calculate the bandwidth requirements of the voice and data at all common points of convergence.

Packet Loss
The third metric is packet loss. Since VoIP is a real-time audio service that uses UDP transport protocols, there is no way to recover lost packets. Packet loss can result in a metallic sound or dropouts in conversations that can be very frustrating to users. The PSTN experiences virtually no loss of digitized voice, but IP networks routinely experience packet loss for a variety of reasons but primarily due to congestion.
To avoid Packet Loss, ensure reliable network connectivity. Planning and deploying a network with no points of congestion and adequate bandwidth throughout the network. Networks deployed as 'Flat' as possible with enough bandwidth to ensure proper data and voice communications result in the best performance. Avoid chaining multiple Layer 2 switches together in a long serial chain, as the higher the up the chain traffic goes the more congested the final up link becomes. Deploy and number of Quality of Service techniques to ensure the best prioritization of Voice and Data. Calculate the bandwidth requirements of the voice and data at all common points of convergence.

3 Quality of Service (QoS)

3.1.1 LAN QoS Options

VLAN and Priority Tags (802.1p/Q)
IEEE 802.1Q (also known as VLAN Tagging) was a project in the IEEE 802 standards process to develop a mechanism to allow multiple bridged networks to transparently share the same physical network link without leakage of information between networks (i.e. trunking). IEEE 802.1Q is also the name of the standard issued by this process, and in common usage the name of the encapsulation protocol used to implement this mechanism over Ethernet networks. IEEE 802.1Q also defines the meaning of a virtual LAN or VLAN with respect to the specific conceptual model underpinning bridging at the MAC layer and to the IEEE 802.1D spanning tree protocol. This protocol allows for individual VLANs to communicate with one another with the use of a layer-3 router. IEEE 802.1p is a standard that provides traffic class expediting and dynamic multicast filtering. Essentially, it provides a mechanism for implementing Quality of Service (QoS) at the MAC (Media Access Control) level. Eight different classes of service are available, expressed through the 3-bit user_priority field in an IEEE 802.1Q header added to the frame. The way traffic is treated when assigned to any particular class is undefined and left to the implementation. The IEEE however has made some broad recommendations. Neither of these IEEE standards extends onto the Internet (WAN). These standards are meant for local networks (LAN) only.

Differentiated Services (DiffServ)
DiffServ or Differentiated Services is a computer networking architecture that specifies a simple, scalable and coarse-grained mechanism for classifying, managing network traffic and providing quality of service (QoS) guarantees on modern IP networks. DiffServ can, for example, be used to provide low-latency, guaranteed service to critical network traffic such as voice or video while providing simple best-effort traffic guarantees to non-critical services such as web traffic or file transfers.

DiffServ operates on the principle of traffic classification, where each data packet is placed into a limited number of traffic classes, rather than differentiating network traffic based on the requirements of an individual flow. Each router on the network is configured to differentiate traffic based on its class. Each traffic class can be managed differently, ensuring preferential treatment for higher-priority traffic on the network. The DiffServ model does not make judgment on what types of traffic should be given priority treatment since that is left up to the network operator. DiffServ simply provides a framework to allow classification and differentiated treatment. DiffServ does recommend a standardized set of traffic classes (discussed below) to make interoperability between different networks and different vendors' equipment simpler. DiffServ does not extend onto the Internet (WAN). This mechanism is meant for local networks (LAN) only.

3.1.2 WAN QoS Options

There are no protocols or standards that can be reliably provide Quality of Service over the Internet (WAN). The Internet is intended as a "Best Effort" network without concerns for reliability and latency.

There are some Internet Service Providers that will offer some Quality of Service capability within their own network. Some use MPLS service, and others may continue to use DiffServ to provide QoS within their own network.

Traffic Shaping (Packet Shaping)

Traffic shaping (also known as "packet shaping") is an attempt to control computer network traffic in order to optimize or guarantee performance, low latency, and/or bandwidth by delaying packets. Traffic shaping deals with concepts of classification, queue disciplines, enforcing policies, congestion management, quality of service (QoS), and fairness.

Traffic shaping provides a mechanism to control the volume of traffic being sent into a network (bandwidth throttling), and the rate at which the traffic is being sent (rate limiting). For this reason, traffic shaping schemes are commonly implemented at the network edges to control traffic entering the network. This control can be accomplished in many ways and for many reasons but traffic shaping always simply consists in delaying packets. Traffic policing is the related practice of packet dropping and packet marking. Traffic shaping can be applied by the traffic source (for example, computer or network card) or by an element in the network.

Ingate has incorporated a Quality of Service Module within our products. With this application the SMB/SME can shape traffic to better utilize their Internet bandwidth. They can prioritize and do bandwidth limitation for different traffic types, based on source or destination IP addresses, type of service, packet size, and TOS/DSCP markup. This ensures the VoIP traffic has primary use of the available bandwidth.

Although traffic shaper technology can control the data in and out of a SMB/SME at the LAN-WAN edge, it does not control or provide a Quality of Service over the Internet.