SD-VOICE Primer

2/24/2016

www.voltesys.com

Graham King PhD
CTO/President
&
Mark Darvill
SVP

3859 Centerview Drive, Ste 500
Chantilly, VA 20151

VoLTE Systems SD-VOICE

Contents

VoLTE Systems SD-VOICE

1. Introduction

The purpose of this document is to describe VoLTE Systems SD-VOICE.

SD -VOICE is a unique virtualized sub-appliance built on a service oriented architecture that can be easily integrated with a Software Definable Wide Area Network (SD-WAN) appliance to provide scalable, policy-based, high quality voice over IP (VoIP) for real time telephony and video collaboration.

SD-VOICE consists of three components; a standards-based VoIP Session Border Controller (SBC), a Public Switched Telephone Network (PSTN) compatible media gateway and a suite of voice-specific IP WAN optimization functions. Since most businesses still rely heavily on communication over the PSTN, a primary function of SD-VOICE is to orchestrate connectivity in the hybrid IT environment, and to ensure seamless operation with PSTN based legacy equipment and services. When integrated with an SD-WAN appliance, SD-VOICE can eliminate the need for a separate Integrated Services Router (ISR), SBC and media gateway in the Branch, reducing complexity, simplifying provisioning and enhancing survivability.

From the user perspective, the combining of SD-VOICE with SD-WAN ensures the highest quality audio for each voice call and greatly improves the consistency and dependability of telephony service over shared public networks such as the Internet.

Sections 2 through 5 start with general descriptions of SD-VOICE and how it fits into the SD-WAN environment. From Section 6 onwards the descriptions become gradually more technical.

2. Background

Voice over IP has been deployed for many years, but the quality of VoIP service can vary considerably, depending on the environment. Within a large enterprise with high-speed network connections at the core and a surplus of bandwidth available to support real time services, standard VoIP can match the best voice quality achievable over the PSTN. If wideband voice is used (i.e. High Definition Voice), the voice quality can easily exceed PSTN "toll quality".

When lower speed WAN connections are used, real time services may be dependent on differential routing, such as over an IP network that supports Multi-Protocol Label Switching (MPLS). Using MPLS, the audio packet stream is given priority over other applications and provided the network resources are available, high quality voice service is assured for the duration of each call.

However the quality of VoIP service suffers when used over economical Internet services that do not offer the protection of MPLS, especially at the edge in the small/medium Branch office where data rates may be limited to one or more relatively low speed connections. The shared resources of the Internet are subject to congestion, which can slow down call setup times and break up audio during a call if packets get delayed or discarded. As a result, Internet telephony has a reputation for inconsistency and poor audio quality.

From the business perspective, reducing dependence on expensive MPLS links by using the Internet can lead to substantial cost savings. However, maintaining high quality telephony is important to maximize employee productivity. SD-WANs offer a potential solution to this problem by providing greater end-to-end visibility and control over bandwidth management. By using a centralized policy that steers traffic over the most appropriate WAN connection, resources can be better matched to the applications running over the network.

According to the Gartner Group, [ref 1] SD-WANs will experience rapid growth as an emerging technology making the Internet a practical option for providing core business services. They estimate that by the end of 2019, 30% of enterprises will use SD-WAN products in their branches, up from less than 1% today.

SD-WANs offer several benefits compared to traditional, router-based WANs. They monitor traffic on the network and determine the best route for each packet based on the application and current network conditions. Each application is prioritized according to its centrally administered Application Network Profile (ANP) and a centralized routing policy. In this way SD-WAN appliances are able to treat one or more independent communications channels as a single channel of higher capacity. This process is known as packet-flow optimization or packet shaping. This distributes the traffic switching functions across the network, reducing support costs and simplifying configuration and provisioning.

SD-WANs have the potential to be especially beneficial to VoIP in organizations with branches that rely on one or more relatively thin connections such as Primary rate T1 (1.544Mbps), Digital Subscriber Lines (DSL), cellular wireless data or satellite links. In this environment, connections carrying multiple voice calls often operate between peer devices at the branches and a regional center. Typically these voice trunks are based on the Session Initiation Protocol (SIP). SIP is a notoriously inefficient protocol that creates unnecessary network loading both in terms of overall data throughput and the number of packets generated.

The introduction of SD-WAN packet-flow optimization can alleviate some of the issues associated with VoIP inconsistency due to packet loss. SD-WANs typically treat SIP packets as having the highest priority, and may also provide additional protection by adding error correction, or simply by duplicating their transmission across different carrier connections. This mitigates the effect of occasional packet loss, but at the expense of using additional bandwidth, which is not ideal (and can easily make matters worse) if the network is already congested. Generally, SD-WANs have no ability to optimize or protect individual voice streams and as a result, SD-WANs alone can only go so far in protecting VoIP call quality.

By contrast, SD-VOICE is a voice-stream optimization technology that can augment or replace existing SIP trunks with more efficient peer connections, often with little or no changes to an existing setup. By combining multiple voice calls into an optimized packet stream an efficiency gain of up to 10:1 can be realized in both the bandwidth used and the number of packets generated. With such a significant reduction in the network resources required, SD-VOICE can greatly reduce packet loss and dramatically improve overall VoIP Quality of Service (QoS). When coupled with SD-WAN policy based routing, voice calls can be further optimized and prioritized based on class of service, becoming more efficiently integrated with high volume applications such as data from cloud applications.

SD-VOICE also inter-operates with external SIP trunks and media servers, and provides gateway functions to the PSTN so that the internal performance advantages are not lost when connecting to outside voice services. The combination of SD-VOICE and SD-WAN therefore ensures the highest quality audio for each voice call and greatly improves the consistency and dependability of VoIP within the enterprise.

In the following sections the capabilities and advantages of SD-VOICE are described, starting with a more detailed description of SD-VOICE, the problems SD-VOICE addresses and how it fits into the SD-WAN environment. Later sections become gradually more technical and highlight some of the unique capabilities of the SD-VOICE system.

3. The Role of SD-VOICE

While many VoIP calls are made directly between IP enabled devices and may never leave the IP infrastructure, it is generally accepted that the vast majority of business users worldwide still rely on the PSTN (including of course access to cellphones) for routine voice communications. SD-VOICE is designed to operate in this hybrid environment of VoIP, PSTN and today's Long Term Evolution (LTE) wireless networks.

A primary role of SD-VOICE when coupled with SD-WAN is to ensure the most dependable, most consistent and highest quality PSTN compatible voice communications across IP networks. Since High Definition Voice used by some VoIP systems is not designed to be compatible with the PSTN it is not discussed further here. However many of the capabilities of the SD-VOICE system can also be applied to High Definition Voice and to video collaboration.

VoLTE systems IP telephony capabilities have been developed over many years with a long heritage that includes voice technology originally developed by Republic Telecom, Netrix, NxNetworks, and NSGDatacom prior to VoLTE Systems. The core components are widely deployed throughout the world, initially used in satellite links by the US Military and for Carrier applications, and in recent years used in numerous cellular wireless and terrestrial applications.

The SD-VOICE sub-appliance has a highly portable Service Oriented Architecture that functions successfully on multiple target operating systems; VxWorks, Windows, eCos, Linux (including Fedora, Redhat, Debian and Ubuntu), among many others. It is structured for easy integration within third party appliances, and comprises three core components:

- A Session Border Controller (SBC) for SIP trunks.
- A Media Gateway (with optional hardware interfaces).
- VoIP Optimization functions.

The VoLTE Systems SBC: controls the signaling and media streams involved in setting up, conducting and tearing down telephone calls. The SBC is necessary to resolve differences between various IP networks and works alongside the VoLTE Systems Media Gateway to interface between standards based VoIP systems, LTE networks, the PSTN, and other Time Division Multiplexed (TDM) systems. The SBC supports trans-coding where necessary and provides security and critical control functions at the ingress and egress of attached Networks. It supports the secure registration of remote users and will inter-operate with, or terminate SIP trunks to a wide range of Soft Switches, Call Managers (CMs), media servers, IP Private Branch Exchanges (IP PBX's), Unified Communications Platforms and external carrier switches.

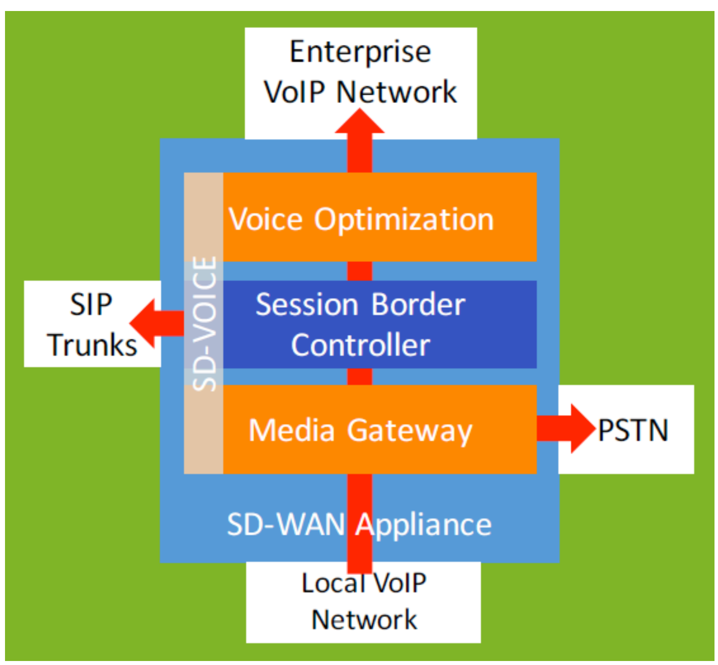
The VoLTE Systems Media Gateway: is an essential component for connecting to non-VoIP enabled devices and legacy networks and supports interfacing to carrier networks worldwide. In conjunction with optional hardware interface modules, the Media Gateway consolidates VoIP and PSTN access, and IP-enables conventional Telephony equipment such as handsets, Point of Sale (POS) terminals and facsimile machines. It provides line equalization (echo cancellation) for direct connection to analog circuits and clock management for digital carrier services, either of which can then be used to access national or international PSTNs. Connection can be through a conventional PBX or Key System, directly to a carrier switch, or over LTE data circuits. Any network connection can be shared as a primary resource or used on demand as backup communications path for business continuity. The Media Gateway offers the potential for SD-WAN appliances to replace incumbent ISRs in a hybrid network environment, reducing complexity, power and cooling costs through device consolidation.

VoLTE Systems VoIP Optimization Functions: are the key to delivering consistent, high quality VoIP. The primary optimization functions include audio compression, packet aggregation and header compression, supported by a host of supplementary capabilities. Many of the SD-VOICE functions are unique in the VoIP market, and can be applied to existing SIP systems as well as other call-based services. SD-VOICE can provide a consistency in call quality not achievable using standard VoIP, especially in a hybrid IT environment. Users normally experience more rapid connect times, fewer dropped calls, and greatly improved audio.

From the business perspective, the optimization functions of SD-VOICE allow cost savings by dramatically reducing the use of network resources, freeing up capacity for other applications.

SD-VOICE also supports a range of carrier class functions, including security and cloud-based configuration and management. These functions can be integrated with the SD-WAN management portal to accept application or topology specific SD-VOICE configuration profiles (ANPs) from the centralized policy controller. Different ANPs can define the alternative configurations necessary to operate in different Branch office environments. These can be distributed via the SD-WAN control plane to the SD-VOICE sub appliances, simplifying installation, provisioning and growth.

Additional information on the capabilities of the Session Border Controller and Media Gateway including the functions listed above can also be found on the VoLTE Systems website: http://www.voltesys.com/



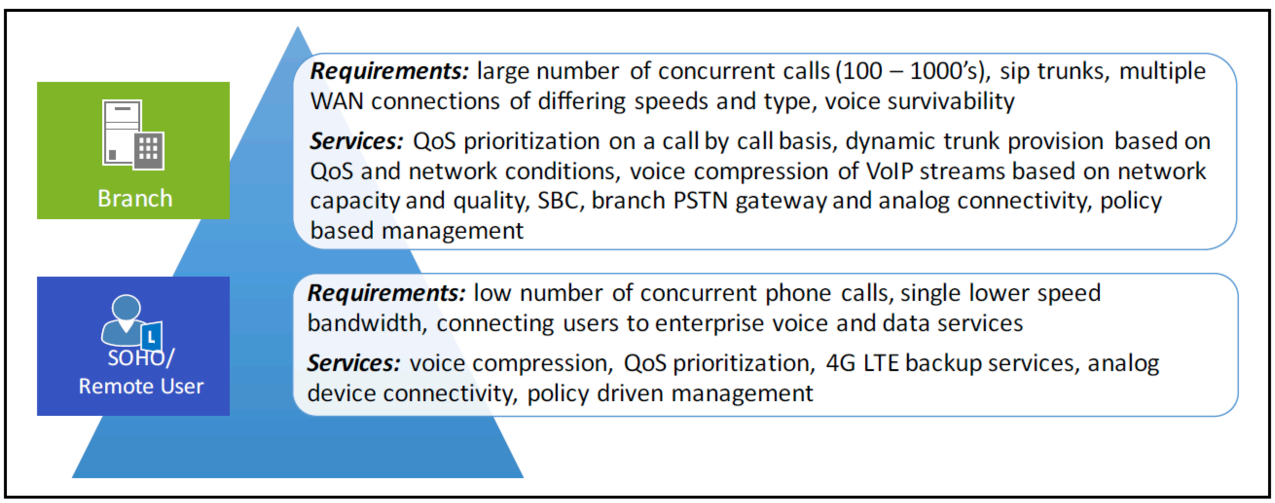
4. Using SD-VOICE in an existing network

It is anticipated that combining SD-VOICE with an SD-WAN appliance, and using centrally managed ANPs to configure the SD-VOICE sub-systems, will allow a range of proven plug and play options to be defined for a wide variety of applications with minimal additional configuration being necessary. These applications range from generic SIP based VoIP trunk optimization, interactions with web based services such as Microsoft Lync or WebRTC, and PSTN bypass functions, to many industry specific solutions (outlined below).

For example, existing SIP trunk optimization can be easily accomplished between Branch offices and a centralized switching center using a simple pre-defined ANP for that topology. Standard G.711 VoIP calls made between conventional VoIP equipment can be optimized by SD-VOICE using both audio compression and packet optimization for a 10:1 resource advantage on the audio packet stream. Other SIP messages may be passed without alteration so that soft switch or CCM functions are not affected.

SD-VOICE can also be used to improve the efficiency of VoIP services when the payload has already been compressed using (typically) G.729A. In this case SD-VOICE packet optimization can provide additional bandwidth savings of up to 3:1 and packet reductions of up to 10:1 or more.

Using a different ANP, remote Branch office users can be directly connected to a centralized PBX (IP or TDM based) by simply connecting to the media gateway using either an IP or analog phone. Each telephone handset (or other device) connected over the IP network to the central PBX can be allocated a unique extension number on the PBX with minimal configuration.



Using this approach each connected device can appear to be part of the PBX system and there is no difference in operation between these remotely connected handsets and handsets directly connected to the PBX. When a user at the Branch lifts a handset the dial tone comes directly from the PBX. Incoming calls from the PSTN can be received directly and outgoing calls are made by dialing the required PBX sequence (normally 8 or 9) to connect to an outside line. Users can be spread throughout the world and all have local PBX telephone numbers, appearing from the outside to all be located in the same place.

Many other capabilities not normally supported by standard SIP based VoIP systems can be used to connect facsimile machines and low cost legacy data devices such as Alarms and EPOS terminals. This includes the ability to connect devices to the Internet of Things (IoT) using LTE M2M services.

ANPs may be defined that meet the requirements of a wide range of turnkey solutions already deployed in verticals that include Energy Consumption, Industrial Control, Air Traffic Control, Healthcare, Banking and Brokerage.

Some of the ancillary functions relevant to these markets include:

- Multiple levels of VoIP call priority
- SIP and IPSec VPN Security
- Time of day routing
- Virtual networking for multiple tenants
- IP fail-over for TDM links
- TDM Clock recovery
- Transparent TDM over IP
- Transmission Level Point controls (TLPs)
- Full and abbreviated dial plan translations
- Delay Dial
- Auto ring down (upon off hook)
- Immediate start
- Permanent connect
- External device control (eg. ATC, Alarms, Military)

Below is a more in detailed description of some of the competitive advantages gained by integrating SD-VOICE into an SD-WAN appliance. This is followed by a technical description of the optimization capabilities that differentiate SD-VOICE from conventional SIP based VoIP systems.

5. Integration of SD-VOICE with SD-WAN

The assumption here is that SD-WAN appliances are at some level applications aware and gather and share information on current network performance with their peers. A telemetry protocol is used to distribute information between appliances, which can be use to prioritize traffic and make optimal routing decisions across multiple IP connections according to an application's ANP and a centralized management policy. A typical application might be to support cloud-based Software as a Service (SaaS) across an Internet connection and an MPLS based service.

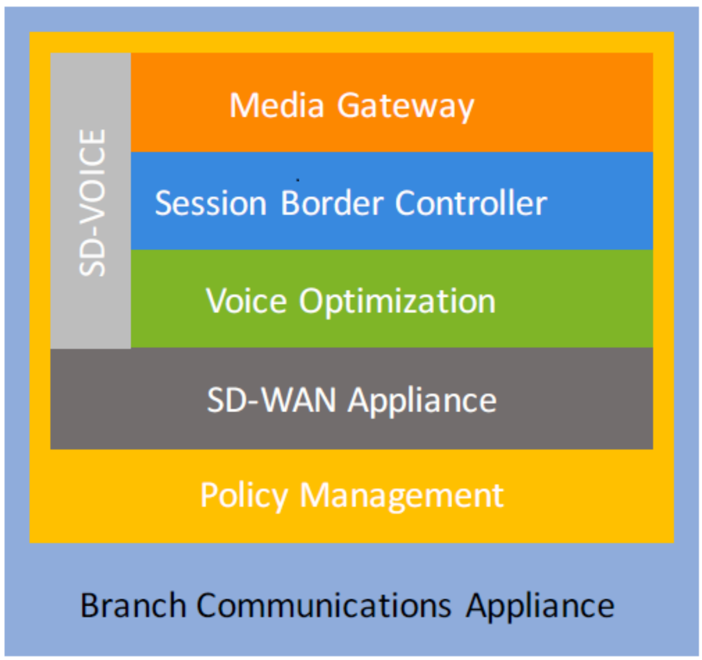
Theoretically, this same information can be used to optimize SIP voice calls. Of course it should be recognized that the requirements for voice are different from other applications running over the network. Most applications are not time critical to the same degree as a real time voice calls, where momentary breaks in service can seriously affect usability. Even in video collaboration the audio component is normally considered more essential and therefore delivery more time critical than the video itself. Also, most applications run relatively asymmetrically in that data transfer in one direction is very much higher than in the other, often by an order of magnitude or more. In networks supporting MPLS, voice is normally given top priority with guaranteed bi-directional bandwidth. However, existing systems do not apply dynamic admission control based on current network conditions and TCP's congestion control algorithm does not take full account of asymetric data flow which may exhibit high capacity in one direction while being bottlenecked in the other.

Ideally, SD-WAN appliances would be capable of using current information to admit, prioritize and optimally route VoIP sessions on a call-by-call basis. However, the connection between peer SIP devices is typically a secure trunk, which often bypasses the SD-WAN appliance completely. Even if a SIP trunk does pass through the SD-WAN appliance, it normally has no control other than to route the trunk packet by packet as a priority, in its entirety. Since there is no opportunity for the SD-WAN to apply individual call admission, optimization or prioritization based on current network conditions, the Branch office is locked into the existing ISR hardware infrastructure and routing policy.

This is not an ideal situation if the objective is to ensure the highest quality telephony service over the Internet. VoIP is often overlooked as having much impact on network resource, being viewed primarily as a very low bandwidth requirement even though it may have the highest priority. Over broadband fiber and cable connections this is mostly true (depending of course on how many simultaneous calls are in progress), but over lower speed Internet connections, especially for thinly connected Branch offices or under emergency situations, VoIP calls can quickly overload resources, seriously affecting the quality of service.

The integration of SD-VOICE with SD-WAN offers the opportunity to address these problems head on and enable a survivable Branch appliance in three ways. Firstly, SD-VOICE can take over control of the SIP trunk from the ISR or other SIP device and apply dynamic admission control and optimization to any existing VoIP communications. Secondly, it can create new voice trunks between peers so that routing and optimization decisions can be exercised independently for every VoIP call as it is established, whether originating from existing VoIP systems, or from the media gateway. Thirdly, SD-VOICE provides the ability to apply levels of prioritization that can be used to manage real-time communications. This may include blocking some calls in order to preserve End user Experience (EuE) if the network is overloaded or in emergency situations, for example, red/black government applications, survivability routing or to ensure capacity is reserved for 911 emergency calls.

The method of operation of SD-VOICE in many ways mirrors that of existing SD-WAN appliances. In common with SD-WANs, SD-VOICE uses telemetry between its peers to collect and share information on link capacity, current loading, route delay, and other factors which, in the absence of an SD-WAN, can be used to make optimal routing decisions. The integration of SD-VOICE with an existing SD-WAN appliance offers the opportunity to share the most appropriate telemetry data. SD-VOICE provides SD-WAN access to an additional level of call-based management detail and status reporting which can be used to extend the centralized, policy-based automation to individual voice streams. This provides a greater granularity of control over the quality of voice services being provided during periods of congestion and for business continuity.



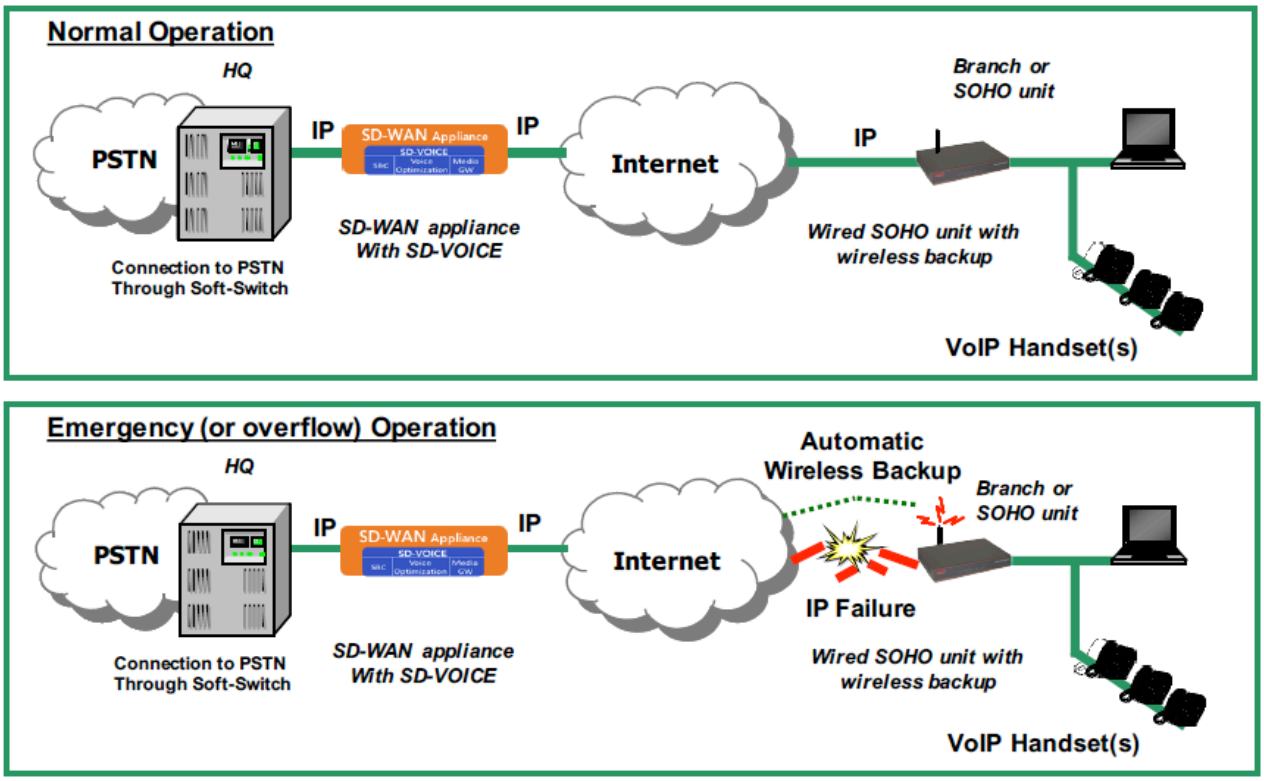
The SD-VOICE telemetry provides a means of supporting ancillary functions, such as emergency access through the media gateway, which can be used to improve resiliency in Branch office scenarios. In conjunction with SD-WAN, survivability routing can be defined that combines current network conditions with SD-VOICE optimization and the ability of the appliance to gateway to outside services on demand, for example to use voice compression over a M2M wireless data link to temporarily bypass a WAN, PBX/Softswitch or MS Lync front end pool failure.

Some of the relevant management functions SD-VOICE telemetry supports include:

- Definable carrier class call detail records
- Automatic node and path discovery
- Self learning, self healing
- Minimal route keep-alive (for dormant/rapid connect)
- On demand 'SIP' trunk establishment

- Remote monitoring
- Redundant clustering

SD-VOICE therefore improves the ability to survive catastrophic events such as a central node crash or complete failure of connected IP networks, by supporting emergency routing through the PSTN or an on-demand service.



Alternatively, SD-VOICE can also be used to bypass a local PSTN line failure using existing IP network connections and the unique TDM over IP fail-over capabilities of the media gateway.

Prior to further describing the optimization techniques in more detail, some of the technical challenges of transmitting VoIP over the Internet are discussed.

6. VoIP over the Internet

There are five closely related problems that VoIP faces compared to conventional telephony. These are network congestion, network delay (latency), jitter, discarded packets, and nodal limitations on packet throughput. All are means by which packet loss can occur in an IP network.

Congestion: When a telephone call is placed over the PSTN a dedicated fixed bandwidth point-to-point connection is guaranteed for the duration of every call. By contrast when a VoIP call is placed over the Internet, the available end-to-end bandwidth changes over time depending on network loading. As a result the bandwidth required to support a continuous audio stream is not guaranteed for the period of a VoIP call. IP Switching nodes cope with momentary network overloaded by storing packets until congestion eases to allow their transmission. At a minimum, the effect of congestion within the network is therefore to increase network latency.

Other congestion factors that extend network latency include the possibility of waiting for some much larger (jumbo) packets to be transmitted, and (if supported) queue jumping by higher priority traffic.

Latency: Every IP switch node in a packet network is a store and forward device. Every packet passing through it is received in its entirety, checked for error, stored for a short period and then transmitted according to its routing. All switches therefore introduce a normally minimal but unavoidable delay in the transfer of an IP packet over the network. An end-to-end delay in transmitting an IP packet over a network is inevitable, and is known as the network Latency.

Although any noticeable latency across the network can be annoying, most people understand and accept the necessity of the delay when talking over a satellite link or even between cell-phones. Users may be willing to compromise on voice quality for convenience, but have also come to accept that significant latency prevents conventional appliances such as fax machines and POS terminals from operating over most IP connections.

Jitter: The change in latency across a network is known as jitter. VoIP receivers always use some buffering to absorb the effects of jitter. In order to accommodate network jitter the buffer depth is normally set to be slightly greater than the longest delay expected across the network. To set the jitter buffer any longer simply adds unnecessary delay to the audio experienced by users. To set the jitter buffer any shorter is to accept that some IP packets containing audio will arrive late, causing an inevitable, (and possibly frequent) break in the audio output. Short breaks in the audio may not be noticeable to the human ear but will terminally corrupt data from fax transmissions or legacy devices such as POS terminals. Long or repeated breaks in audio also render the VoIP quality unacceptable to users as well.

Jitter is a major problem for VoIP because the only way to eliminate the effect of jitter completely is to use a buffer length at the receiving end of the VoIP link that exceeds the maximum potential jitter of the IP network. This introduces a fixed delay into the audio, which if longer than a few hundred milliseconds becomes annoying to users. For all practical purposes any IP packet that arrives outside the jitter buffer window of the system has to be treated as a lost packet, even though it may arrive a moment later.

Discarded Packets: IP networks routinely discard or 'lose' IP packets. Everyone who uses an Internet browser knows the frequency with which pages have to be reloaded or stop half completed for unknown reasons. Every switching node in an IP network can discard packets when there is congestion and local buffer space is becoming exhausted. Some devices use a queue size management algorithm that starts to discard packets in anticipation of increasing congestion. Also, any IP packet received with an error is automatically discarded. There is no error protection or packet retransmission in standard VoIP so there is no automatic recovery from missing or corrupted packets. Discarded or lost packets therefore result in a break in audio at the receiving end of the link unless other actions (such as forward error correction or the multiple transmission of each packet) are taken to protect the integrity of the packet stream.

Limited Packet Throughput: A sometimes-overlooked requirement of network nodes is to meet a minimum IP packet throughput. Some smaller devices have limitations on packet throughput that can affect VoIP systems, especially if a large number of simultaneous SIP calls are in progress.

As a simple example, a typical SIP voice call using a 5ms sample period can generate up to 200 IP packets per second (200pps) in each direction (400pps in total per voice call). An IP switch supporting just 10 simultaneous SIP voice calls can therefore be processing upwards of 4,000pps which can easily exceed the capacity of some smaller network devices, for example, those used in cellular wireless and satellite systems.

All the above limitations must be taken into account for a network to support VoIP quality comparable with the best PSTN call. Clearly there are tradeoffs between these parameters, such as the size of a jitter buffer and the overall network latency that results when initiating a VoIP session. Typically such parameters are configured well in advance and do not take into account current network conditions.

The integration of SD-VOICE with SD-WAN offers the opportunity to create manageable policies for determining the best routing decisions and set operational parameters based on measured network loading, delay and stability. The route with the lowest latency may be the route with the highest packet loss and therefore less desirable for voice. The optimal jitter buffer size might be assigned on a call basis taking into account the current network conditions and jitter profile. The decision to optimize using compression might be made based on a selected combination of factors. For example based on the number of calls in progress, time of day and the amount of free network resource.

Regardless of the routing and optimization policies adopted ultimately all VoIP systems still have to cope with some lost, late or discarded packets.

7. Coping with Missing Packets

Since any connection across the Internet is subject to over subscription, IP packet delay always has the potential of being extended beyond the size of the jitter buffers of the VoIP system. At the extreme, this results in broken or garbled speech and the possibility of dropped calls.

Once the delay threshold required to ensure continuous voice communications is crossed due to missing or late arrival of IP packets, the only option is to cut the audio completely or cover for the dead time some other way. Of the many techniques used to do this, it is common to attempt to approximate the missing audio somehow, (ie. for all practical purposes, guess). More sophisticated systems such as SD-VOICE use elegant methods to connect across missing speech samples, for example, inserting re-processed voice samples to substitute for lost packets in order to keep the audio appear to be flowing normally. Provided there are only isolated packet losses the user is not normally aware that such refinements have even taken place, and abrupt transitions that can be jarring to the listener have been avoided. This does not however protect facsimile or modem transmissions, which typically require no breaks in transmission at all to survive.

While isolated, short gaps of up to a few tens of milliseconds may not be very noticeable, failures can extend well beyond this, for example if network re-routing is required mid-call. To make matters worse, recovery of the packet stream may not be clean. As a result, even when the packet stream is re-established it may not be unusual to hear the audio output restart several times in short succession. This can lead to one of several commonly experienced corruptions to the audio such as voice rapidly cutting in and out, or 'motorboating' due to a rapidly repeated phrase. SD-VOICE is designed to minimize these annoying effects by using a more elegant means of covering for missing voice samples than the repeated use of the same substitute voice sample.

Another commonly used technique to cover for a period of missing audio is to insert complete silence – which for a few moments can sound like the call has dropped completely. During true periods of silence when audio is not flowing, for example during call setup, SD-VOICE uses low volume “comfort noise” which is generally considered to be more pleasing to the listener than absolute silence.

The only realistic solution to the problem of coping with lost or delayed VoIP packets is to minimize the chance of this occurring. Once VoIP packets are delayed beyond the jitter buffer window, whether lost or not, the opportunity to make any significant contribution to the quality of the received audio stream has mostly passed.

In contrast to other systems that use multiple transmissions or forward error correction as a means of improving the timely delivery of voice packets, SD-VOICE does not increase the traffic on the network. SD-VOICE uses the techniques already introduced, ie., policy based routing, audio compression, packet aggregation and packet optimization to improve voice QoS. In conjunction with SD-WAN, these functions can have a significant positive effect on all aspects of call quality including the time to connect, audio fidelity and the ability to maintain long duration calls, while also freeing up resources for other applications.

8. Comparison of the PSTN with Standard SIP-based VoIP Calls

Theoretically, PSTN voice calls and standard, uncompressed SIP based VoIP calls should sound identical because they use exactly the same encoding scheme.

Conventional TDM telephone systems worldwide convert analog voice calls into a digital data stream for transmission. The encoding standard used is G.711, which outputs a constant bit rate of 64Kbps. This conversion is normally carried out at the local telephone exchange and is based on a coding method known as Pulse Code Modulation (PCM). After transmission across the telephone network the PCM data stream is converted back to an analog signal at the receiving end of the link. As with all coding techniques the output approximates (in this case very closely) the analog input signal from the telephone device. For all practical purposes the human ear cannot tell the difference between the original analog signal and the regenerated PCM version. Since a G.711, PCM encoded signal supports the highest level of fidelity achievable on a PSTN voice line, fax machines and dial data modems have been optimized to work with this protocol.

Most Carrier and Cable based VoIP solutions also encode the audio from a local handset according to the G.711 standard and insert the resulting audio stream directly into the IP packet network without further processing. This approach should, in principal at least, provide the highest quality PSTN compatible VoIP and minimize the complexity and cost of end-user equipment. It also makes transitions from a VoIP network to the standard TDM network relatively cost effective and straightforward.

In order to relay the audio over an IP packet network, standard VoIP access devices divide each G.711 audio stream into a sequence of samples of equal time period, insert them into IP packets and transmit the full 64Kbps of data over the IP network. In a perfect world the resulting VoIP calls should be indistinguishable in quality from standard PSTN based telephone calls. (Cellphone calls by comparison compress the audio further so they cannot even in principal match the highest quality of standard PSTN calls). Large Carriers and other mainstream service providers that control dedicated bandwidth right to the consumer can ensure that the required bandwidth is available at all times and can for all practical purposes achieve the goals of VoIP transparency with TDM services.

However, dropped packets, latency and jitter over shared public IP networks most often combine to prevent the ideal VoIP conditions from being realized in practice. Even when VoIP voice calls sound almost perfect to the human ear, (especially when compared to cell-phone calls), facsimile and POS calls frequently fail completely due to the above effects. Packet duplication and forward error correction can improve the situation when there is bandwidth to spare, but in the Branch environment, increasing the load on an already overloaded network connection is probably not the best solution. In reality, fewer, smaller packets nearly always have a better chance of making it across a congested IP network. One common technique used to improve VoIP service capability is to use IP packet header compression. More advanced products, such as SD-VOICE, also use IP packet aggregation when possible. However, the most effective process used to decrease the total amount of VoIP data being sent is to use voice compression on the G.711 payload itself.

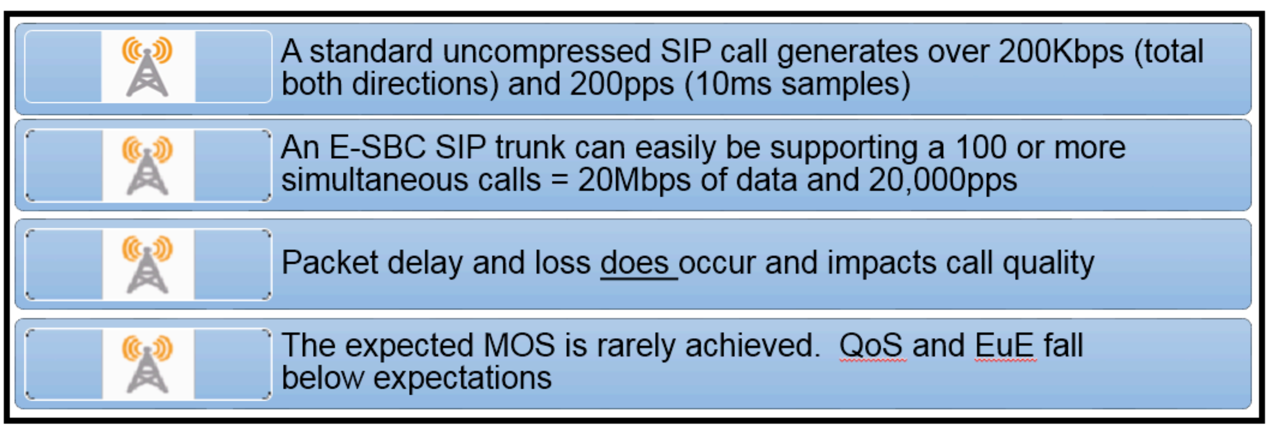
9. The benefits of Compression

In order to best appreciate the advantage and limitations of VoIP voice compression it is useful to initially look at the bandwidth required to transmit uncompressed VoIP calls over an IP network, and to examine the effect this has on the capacity of conventional fixed rate T1 connection.

If we consider an uncompressed (PCM) voice call at 64Kkps and convert this to a standard VoIP call using a 10ms sample period, the raw bandwidth requirement is increased to 107Kbps in each direction.

[Note: 64Kbps of audio creates 100 x 10ms voice samples per second, where each voice sample comprises 80 bytes of audio (PCM) data. (80 bytes x 8 bits x 100 samples = 64,000bits). Added to each 80 byte sample is the overhead of a TCP/IP packet which including the Ethernet frame typically comprises another 54 bytes of data. The total size of the packet transmitted every 10ms is therefore 134 Bytes, which equates to a bandwidth requirement of 107Kbps].

The impact of converting existing service platforms to standard VoIP without compression can therefore be dramatic. The raw bandwidth required to support a standard VoIP voice call is 67% greater than it is over a standard PSTN network. But it gets worse. Whereas the PSTN uses Time Division Multiplexing (TDM) to allocate a fixed bandwidth timeslot for each call, IP networks use contention to allocate network resource before the transmission of every packet. As a result, not only is 67% more of raw network bandwidth required to support each VoIP call, but the contention mechanism adds even more overhead to the system. In practice any IP network loading beyond around 80% of maximum capacity significantly impedes access and congestion starts to interrupt traffic flow. Overall this means that a typical VoIP call requires more than twice the network bandwidth of a standard G.711 PSTN call to be successful.



As a result, whereas a single TDM, T1 (or Primary rate ISDN) line can accommodate the equivalent of 24 simultaneous G.711 voice channels, the identical IP bandwidth can only support 11 or 12 simultaneous VoIP calls (barely 50% of the capacity of the original TDM line) prior to the onset of packet delivery failures due to congestion.

Interestingly, if the sample period of 10ms is doubled to 20ms (halving the number of packets per call), once congestion is taken into account the capacity of standard VoIP on a T1-based IP line barely increases to just 14 calls – still only 63% of the capacity of the original TDM line.

[Note: for uncompressed VoIP calls using a 20ms sample period the resulting packet size increases to 214 Bytes at 50pps, which equates to 85.6Kbps for each VoIP call.]

The basic problem is that the IP packet overhead required to ensure a continuous packet stream is significant compared to the requirements of the TDM network. The implementation of additional services such as encryption using a Virtual Private Network (VPN) can make this even worse. As a result, many small carriers and enterprises have experience problems when converting from TDM voice to standard VoIP calls using existing network bandwidth. Satellite networks in particular cannot cost effectively absorb the additional overhead required by IP telephony and almost always use voice compression to help alleviate this problem.

10. SD-VOICE Compression Options and MOS

Some advanced VoIP equipment and handsets provide the option to compress the standard PCM encoded audio. There are many compression algorithms available but the most common VoIP compression techniques use a version of a complex Algebraic Code Excited Linear Predictive Algorithm (ACELP), of which there are many variants. The most common ACELP standard adopted over satellite links and in some VoIP handsets is G.729A, which reduces the audio digital data rate from 64Kbps to just 8Kbps.

Of course, any algorithm that achieves an 8:1 compression ratio is bound to reduce the quality of the audio signal compared to the original PCM stream. The generally accepted method of defining the audio quality over telephone networks is according to a scale known as the Mean Opinion Score (MOS). The MOS is based on subjective analysis in a controlled environment and graded between 0 and 5.0, where the perfect MOS score is 5.0. A score of 3.0 is considered slightly annoying. PCM is generally scored at MOS 4.1, which is described as perceptible but not annoying. G.729A has been a very successful standard due to the surprisingly high level of audio fidelity that is retained given the high compression ratio. G.729A is normally scored at MOS 3.7. Alternative versions of Linear Predictive Algorithms including SD-VOICE ACELP score as high as MOS 3.9.

SD-VOICE calls may be compressed using either G.729A or VoLTE Systems unique advanced ACELP, which also generates an 8Kbps compressed audio stream. SD-VOICE ACELP has been independently tested and verified to meet or exceed MOS 3.9. It was originally developed for military use and has been widely chosen for deployment in US Military satellite systems over the past 20 years due to its low bandwidth and very high voice quality. Other, lower bandwidth options are also supported by SD-VOICE but are rarely used due to the exceptional audio quality obtained using the 8Kbps ACELP algorithm.

However, placing too much emphasis on MOS alone can be a very misleading. MOS is a subjective assessment of the voice quality under laboratory conditions that cannot fully take into account actual network delay or effects due to covering for missing packets. In practice the advantage gained by moving from uncompressed G.711 PCM voice to compressed voice at 8Kbps more then offsets the small loss of MOS that results from the compression itself. In other words, once some packet loss is being experienced moving from uncompressed to compressed voice can make a noticeable improvement in voice quality, which is the opposite of what the MOS would suggest. Even under conditions where packet loss is unlikely, ACELP voice compression is virtually indistinguishable from uncompressed VoIP and offers substantial advantages by freeing up resource for other applications.

The MOS of a voice call relates to the quality of the audio as it sounds to the human ear. Accordingly, the commonly used compression algorithms (including G.729A and SD-VOICE ACELP) have been tuned to minimize the audible impact of the compression on human speech. Such algorithms do not attempt to retain other components of the signal that are not critical to voice recognition, such as the specific frequency and phase relationships used by analog modems and facsimile machines. As a result neither of these will operate over a VoIP channel that is subjected to such aggressive compression, even though there may be minimal discernable difference in the quality of the audio to the human ear. Standard FSK tones (telephone handset key presses) can also be distorted to the point they are not decodable by the receiver after decompression. The primary distortion is introduced by the compression and decompression process, further aggravated by any effects of network induced latency and packet loss.

Support for the local regeneration of FSK tones is essential if standard telephone key presses are to cleanly transit a VoIP network, for example to dial an extension number after connecting to the automatic attendant of a PBX or soft switch system. This is one of the functions provided by the subsystems of SD-VOICE. Other specialized functions provided by SD-VOICE discussed below include support for facsimile machines and modem dependent devices such as utility meters and alarm systems.

The benefits of compression therefore have to be weighed against its disadvantages, which are primarily related to non-voice applications. SD-VOICE can determine while a VoIP session is being established if it is a facsimile call, in which case the mode of operation is changed to support Fax over IP.

By itself, voice compression using G.729A or using SD-VOICE's high quality ACELP can provide a useful 8:1 reduction in the audio payload bandwidth used, which in turn helps to improve the likelihood that VoIP packets are delivered successfully without interruption. However the real life bandwidth savings are not as substantial as they sound initially once the IP overhead is taken into account.

11. Compressed VoIP Call Bandwidth

Using the same methods of calculation as above, if the audio payload of a VoIP call is compressed to 8Kbps the equivalent calculation yields a packet size of 64 bytes, and a data rate of 51.2Kbps.

[Note: 8Kbps of compressed audio divided into 100 voice samples per second, where each voice sample comprises 10 bytes of compressed data. (10 bytes x 8 bits x 100 samples = 8,000bits). Added to each 10 byte sample is the overhead of a TCP/IP packet which typically comprises another 54 bytes of data. The total size of the packet transmitted every 10ms is therefore 64 Bytes, which equates to a bandwidth requirement of 51.2Kbps per call.]

Therefore a compression ratio of 8:1 on the voice payload yields just over 50% of bandwidth savings on the channel itself when compared with uncompressed (G.711) VoIP. Often, for this reason compressed VoIP calls combine two sample periods and use a 20ms payload period, which adds an additional delay of 10ms, but yields a more acceptable 29.6Kbps bandwidth requirement for each call. Even so, less than 30% of the required bandwidth is being used to carry the 8Kbps voice payload.

It is clear that the IP overhead has become very significant compared to the voice payload. Additionally, each voice call is still generating 100+ packets per second (50pps in each direction for 20ms samples) so there is still the potential for a high number of packets generated by a VoIP switch or call manager to create bottlenecks in some networks. Nevertheless independent VoIP streams using G.729A compression and a 20ms payload period are a common VoIP combination used by SIP systems.

Very few products on the market are designed to improve upon this combination. SD-VOICE ACELP by comparison is based on a unique 18ms sample period, which due to the strength of the algorithm, simultaneously improves the audio quality while also reducing the core processing resources required, ie. it uses one slightly more complex ACELP 'crunch' every 18ms rather than two standard ACELP 'crunches' every 20ms.

While voice compression can make a major contribution to bandwidth utilization, the additional application of SD-VOICE packet optimization can improve upon this even more, and create an overall 10:1 resource advantage. Furthermore the SD-VOICE packet optimization functions can be applied to standards based G.711 or G.729A VoIP calls as well as other real time streams such as video.

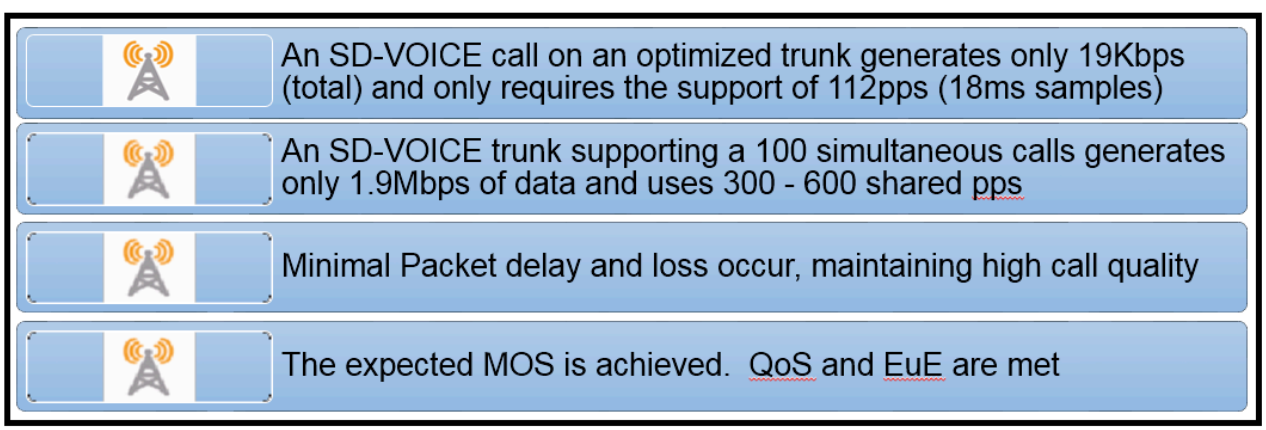
12. Packet Optimization

SD-VOICE optimizes the bandwidth used for telephony services over public (Internet) or private IP connections using a combination of voice compression and packet optimization techniques. The processes used in SD-VOICE extend far beyond those typically used to optimize VoIP services and are controlled by the SD-VOICE telemetry protocol known an SFTM (Switched Frame Transfer Mode).

SFTM is VoLTE Systems advanced networking protocol used between SD-VOICE appliances. SFTM is more efficient than standard IP based protocols (such as the Real-time Transport Protocol - RTP) for several reasons. Firstly, SFTM uses basic IP datagrams (UDP packets) rather than full IP packets whenever possible in (for example) the enterprise environment. Secondly, the SFTM header information itself is shorter than either standard RTP or compressed RTP. Thirdly and most significantly, multiple packets to the same destination are automatically multiplexed into a single packet to dramatically reduce the total number of packets are transmitted between nodes, resulting in a major reduction in the overall bandwidth used.

For example, using these techniques, even a single compressed VoIP call over an IP network can be reduced from 29.6Kbps (see above) to just 15.2 Kbps per call. When just three calls are active the bandwidth drops to only 11 Kbps per call. When ten calls are active the bandwidth used is just 9.5 Kbps per call. Most significantly, the audio quality is not impacted. In fact the overall quality of the call is normally improved compared with the standard SIP trunk alternative, since fewer packets are likely to be delayed or dropped by the network.

The practical effect of such a dramatic bandwidth saving cannot be over emphasized. For example as discussed above, a T1 rate TDM line than supports the equivalent of 24 PSTN calls, when converted to IP only supports somewhere in the range 11 to 14 standard (uncompressed or compressed) SIP VoIP calls. However, using SD-VOICE SFTM technology the same T1 bandwidth can comfortably support well over 100 toll quality VoIP calls, with capacity to spare.

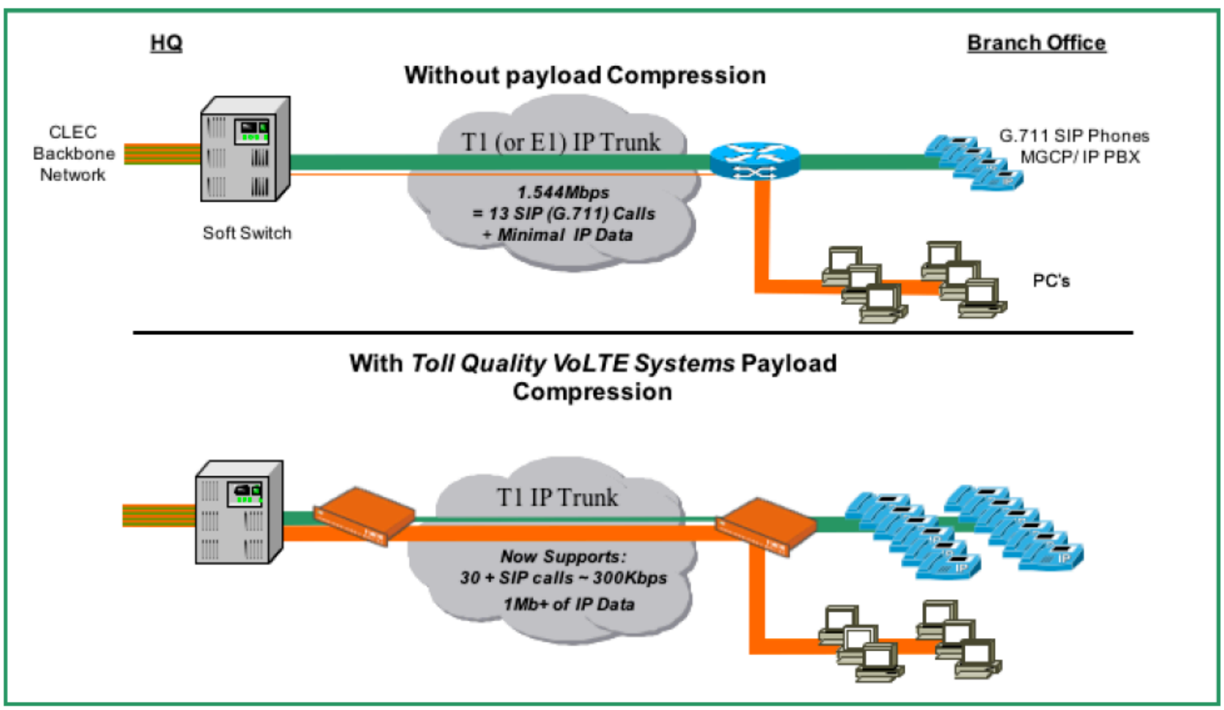


Furthermore the total number of packets required to support 100 simultaneous voice calls can be as few as 400pps (200pps in each direction).

This compares with 10,000pps for standards based compressed VoIP (20ms sample period = 2 x 50pps x 100 calls = 10,000pps). As a practical consideration, regardless of the number of simultaneous calls being supported, a typical setting for a SD-VOICE trunk of this capacity would be in the region of 300pps to 600pps. This minimizes latency and keeps both packet size and packet throughput within reasonable limits.

As a practical example shown below, in a typical small office environment supporting just 30 voice calls, SD-VOICE optimization releases well over 1Mbps of bandwidth on a T1 link for other applications.

It should be noted that these bandwidth and packet throughput savings are not obtained by using aggressive methods such as silence suppression or ultra low bandwidth voice compression, both of which are additional (but less frquently used) options supported by SD-VOICE and briefly discussed below.



13. Additional Optimization Techniques

In addition to packet optimization, a further technique available on SD-VOICE is silence suppression. Silence suppression is a method whereby the (compressed or uncompressed) audio during periods of silence is not transmitted over the link. Instead, short commands are sent between appliances to indicate the period of silence, and the output audio at the receiving and of the link is stopped during that period. An audio energy level threshold is set to trigger periods of silence suppression on and off at the transmitting end of the link.

Depending how it is implemented, silence suppression can either be unobtrusive or be very noticeable to users with the potential for abrupt cuts in the audio at the beginning and end of speech segments - often described as clipping. As a result silence suppression is generally avoided and only used when bandwidth comes at a high premium.

The effectiveness (and intrusiveness) of the silence suppression mostly depends at what energy level the threshold is set. If the energy threshold is set aggressively high the audio payload bandwidth savings can exceed 50% (since theoretically only one person is talking at a time). However most users find this level of silence suppression leads to very intrusive clipping. More realistic savings are closer to 30% when the energy threshold is lowered to a less intrusive level, although some slight clipping of the audio is still noticeable to most people.

Another of the capabilities of SD-VOICE is to add in "comfort noise" during periods of silence when silence suppression is enabled. Comfort noise level can be set to duplicate the background channel noise normally heard on a land-line telephone when there is no audio activity, effectively making the periods of active silence suppression less noticeable to the listener. The ability of SD-VOICE to combine comfort noise with silence suppression can be used to advantage on bandwidth-restricted links where it represents a much better alternative than implementing more aggressive voice compression. With SD-VOICE optimization, silence suppression can yield an overall bandwidth optimization advantage of up to 16:1 while still retaining PSTN quality audio and minimal clipping.

Whereas the voice quality of calls using 8Kbps SD-VOICE ACELP is indistinguishable by most people from conventional PSTN telephone calls, higher rates of compression are possible but result in a significant reduction in voice quality. Alternative rates of ACELP compression provided by SD-VOICE (but rarely used) are at 7.4Kbps 5.5Kbps and 4.8Kbps.

SD-VOICE also offers an enhanced version of 8Kbps G.729A compression which is compatible with the open standard and provides a slight but noticeable improvement in audio quality when compared directly with VoIP equipment that uses standards based G.729A. This option may be implemented if there is a requirement to interface directly to external VoIP equipment that uses G.729A compression.

14. Coping with Latency and Jitter – Voice Calls

Regardless of the compression technique used, latency and jitter are inevitable when using VoIP connections. In very fast private networks and when used by carriers for local access, latency and jitter may be minimal and the effects inconsequential, but over public IP or Internet connections they vary greatly and can be significant. Satellite based systems of course cannot avoid a 250ms trip delay although they may exhibit minimal jitter. 3G cellular data wireless data connections often exhibit delays measured in many hundreds of milliseconds, sometimes up to a second, and jitter that is not much less. 4G cellular data is normally far less prone to jitter or delay than 3G network data but both are still present.

The primary way to accommodate jitter is to use a jitter buffer that exceeds the maximum expected variation in delay over the network. However, if the delay is maintained at much more than a few hundred milliseconds it becomes very noticeable and eventually unacceptable, even for voice calls. Some early VoIP systems used a variable delay where the jitter buffer effectively expanded to accommodate the maximum delay experienced on a call so far. Unfortunately this resulted in the delay gradually growing during the period of a call, often to a point where the connection was no longer comfortable to use, at which time the only solution was to hang up and redial. The best VoIP systems, such as SD-VOICE, now keep the delay constant and cover for missing or late packets using unobtrusive methods discussed above.

15. Coping with Latency and Jitter – Fax/Modem Calls

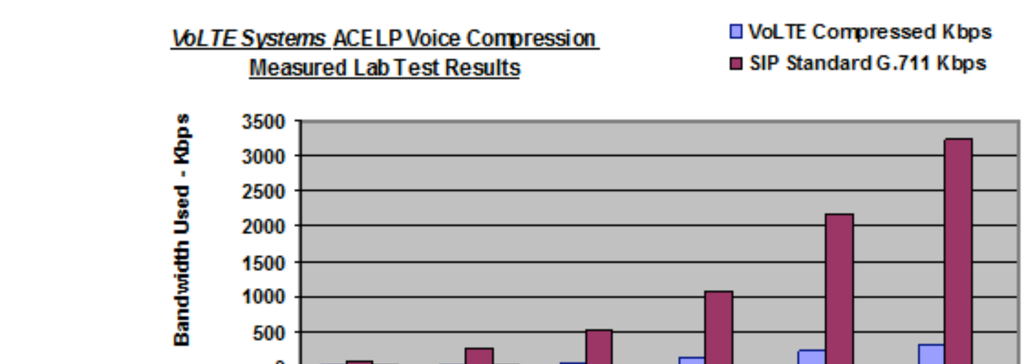
Where real-time fax and modems calls are concerned, lost packets can easily create a terminal problem at the receiving end of the link. Special handling is required for each type of transmission. VoLTE Systems patented techniques used in SD-VOICE include the capability to identify communication sequences that might indicate a fax, modem or other type of tone based 'data' call is taking place and to operate in a different mode from that point in the call onwards. The basic processes involve a combination of techniques that include tone detection, digital transmission of action events, decoding standard modem data and the ability to detect and absorb repeated messages that occur due to timeouts and/or lost packets during non-critical portions of the communication. There are a number of different modes that can be used depending on the application and modem protocols being used.

Real time fax support in SD-VOICE uses some of the techniques employed by the T.38 standard, which are enhanced by VoLTE Systems functions to accommodate the effect of extended jitter and delay. As a result, typical G.711 or T.38 fax calls that fail over a VoIP network can be successfully made using the SD-VOICE appliance. This includes the reliable support of fax calls over 3G cellular wireless data networks that might incur delays and jitter of up to 1 second or more. The facsimile component of SD-VOICE has been successfully tested and approved by Sprint and Verizon Wireless as able to successfully transmit faxes over the 3G cellular data network, and has been routinely used to send facsimiles over both single and double hop satellite networks.

Dial modem calls and other FSK tone based communications (such as alarm panels or simple touch pad key presses) can also be supported in real time rather than using the more frequent and less desirable technique of store and forward, also sometimes known as modem relay. As already discussed SD-VOICE uses a technique that recognizes and regenerate tone sequences at the far end of the link, which avoids distortion due to compression and reduces the negative effects of lost packets. This latter capability is enhanced with patented software techniques that manipulate certain parameters in the transmission and response sequence in order to improve dependability of FSK modem communication over IP data networks.

16. SD-VOICE ACELP Performance graphs

The graph below shows typical lab results comparing the bandwidth advantage of using SD-VOICE ACELP (18ms sample periods) with standard uncompressed VoIP (20ms sample periods) over VoIP trunks.



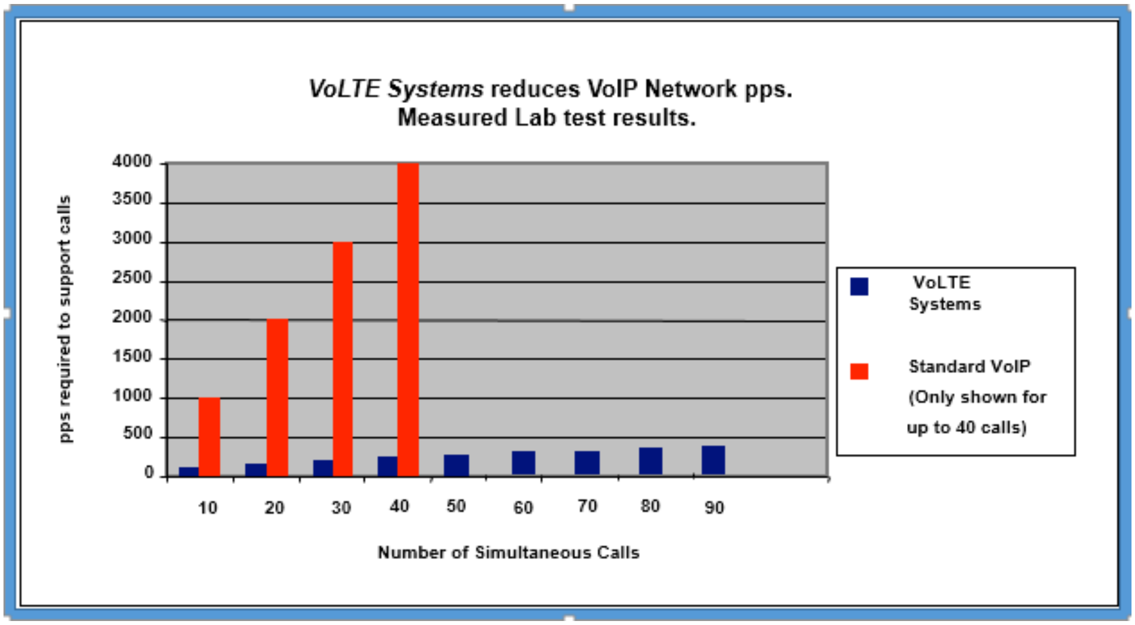
Number of concurrent calls	1	3	6	12	24	36
VoLTE Compressed Kbps	15.3	33	61	114	226	337
SIP Standard G.711 Kbps	90	270	540	1,080	2,160	3,240
VoLTE Kbps per call	15.3	11.0	10.2	9.5	9.4	9.4

20ms sample period, Toll Quality Voice with no Silence Suppression

The exact performance gain is dependent on the sample periods being compared. The graph above shows that for 12 or more simultaneous calls the total bandwidth used by SD-VOICE compared with standards based G.711 SIP calls using a 20ms sample period is approximately 9:1. By contrast, if a 10ms SIP sample period is used the advantage is over 11:1 (not shown).

The second graph shows typical lab results comparing the savings in packet throughput using SD-VOICE packet optimization compared to standard VoIP (20ms sample periods) over SIP trunks. SD-VOICE packet optimization can be applied to any voice packet stream regardless of compression type.

Using SD-VOICE the number of calls supported by a single packet stream can be defined in a number of different ways. Although 10:1 is used as an example in the text, typically for highly utilized trunks, a packet reduction of 20:1 or more can be easily achieved.



17. Summary

SD -VOICE is a virtualized sub-appliance that offers greater throughput, higher voice quality and more capability than conventional VoIP systems and equipment. It has a proven and widely deployed services oriented architecture that can be easily integrated with a Software Definable Wide Area Network (SD-WAN) to provide scalable, centralized ANP (policy-based) control and high quality IP telephony, particularly in the Branch office environment.

The combining of SD-VOICE with SD-WAN provides for greater granularity of control over the quality of the voice service being provided. From the user perspective this ensures the highest quality audio for each voice call in a hybrid IT environment, and greatly improves the consistency and dependability of telephony service both within the enterprise and across shared public services. Users normally experience more rapid connect times, fewer dropped calls, and greatly improved audio.

From the business perspective, the optimization functions of SD-VOICE allow cost savings by dramatically reducing the use of network resources, freeing up capacity for other applications. When integrated with an SD-WAN appliance, SD-VOICE can eliminate the need for a separate Integrated Services Router in the Branch, reducing complexity, simplifying provisioning and enhancing survivability.

The SD-VOICE sub appliance consists of three components; a standards-based VoIP Session Border Controller (SBC), a Public Switched Telephone Network (PSTN) compatible media gateway and a suite of voice-specific IP WAN optimization functions. Since most businesses still rely heavily on communication over the PSTN, a primary function of SD-VOICE is to orchestrate connectivity in the hybrid IT environment, and to ensure seamless operation with PSTN based legacy equipment and services.

SD-VOICE optimization functions are the key to delivering consistent, high quality VoIP. In addition to packet shaping the primary optimization functions include audio compression, packet aggregation and header compression. Many of the SD-VOICE functions are unique in the VoIP market, and can be applied to existing SIP systems as well as other call-based services.

SD-VOICE operates over connections where other VoIP equipment repeatedly fails such as cellular wireless 3G or 4G data connections. Whereas standard VoIP calls require much more bandwidth to support than standard PSTN calls, (typically around 100Kbps compared to 64Kbps), SD-VOICE can reduce this to under 9.5Kbps per call including the TCP/IP and Ethernet packet overhead, which can dramatically improve call quality over congested networks.

The practical effect of such a dramatic bandwidth saving is substantial. For example a T1 rate TDM line than supports the equivalent of 24 PSTN calls, when converted to IP only supports somewhere in the range 11 to 14 standard (uncompressed or compressed) SIP VoIP calls. However, using SD-VOICE SFTM technology the same T1 bandwidth can comfortably support well over 100 toll quality VoIP calls, or support 30 voice calls (more than a standard TDM T1 trunk) and still leave over 1Mbps of capacity available for other applications.

The US Military, ATT and many other global organizations and carriers have been using SD-VOICE based solutions for over 15 years.

Additional functions that extend beyond the normal capabilities of VoIP have been certified by Verizon and Sprint for use over cellular wireless M2M Data services. These include patented techniques that support facsimile machines and the connection of devices containing analog modems to the 'Internet of Things'. Some of the vertical markets served include brokerage (dealer boards), air traffic control, telehealth, and metering among others.

Reference

Ref 1: Market Guide for Software Defined WAN. Gartner Group Report, 1st December 2015.