Strategies for Delivering Reliable Call Quality in Cloud-Based Business Phone Systems

Voice over Internet Protocol (VoIP) has steadily grown in popularity primarily because of its cost-effectiveness over the traditional public-switched telephone network (PSTN). VoIP has moved from being a pure voice replacement for PSTN to a value-added service that seamlessly integrates voice, video, and data.

Your business phone system is a lifeline to your customers, suppliers, and contacts. It should serve your employees in the office and those working remotely. Ideally, it should also support faxing, conferencing, and video meetings. Above all, your business phone system needs to be reliable and provide a consistent Quality of Service (QoS).

In this paper we discuss:

- Essential considerations for delivering reliable VoIP to your enterprise.
- Factors affecting network Quality of Service.
- Technologies for improving end-to-end QoS over the Internet.
- Recommendations you can use to evaluate VoIP vendors.
Factors Affecting QoS

Every cell phone user has experienced spotty coverage and inconsistent reliability. You’ve come to expect this and you move to a better location. A business, however, cannot move about in pursuit of a better, more reliable connection. Your business depends on your cloud vendor to provide consistent, high-quality coverage at your present location.

The following factors can profoundly impact Quality of Service over a network.

Network Delay

Two problems can arise from end-to-end delay in a voice network: echo and talker overlap. Round trip delay in a voice network should be less than 50 ms to avoid echo problems. Since VoIP typically has longer delays, echo control and echo cancellation methods must be employed.\(^1\)

Talker overlap (one caller interfering with another caller’s speech) will be significant if the one-way delay in a network is greater than 250 ms. Network delays compound as the voice packet travels through the network. Employing a fast CODEC such as G.279 CS-ACLEP helps take care of accumulation and processing delays, while network delay can be minimized by a network design consisting of fewer hops and faster Layer 3 switching devices, such as MPLS (Multi-Protocol Label Switching) systems and ATM switches.

Jitter

Jitter is the variation in the arrival time of data packets, which results in gaps between the packets. The result of high jitter (over 50 ms) is speech that sounds jerky. Severe jitter can cause voice data packets to arrive out of order, which then causes sounds to be jumbled. Jitter is caused by the variable transmission delay over the network.

Removing jitter requires collecting packets in buffers and holding them long enough to allow the slowest packets that arrive in time to be played in correct sequence. Jitter buffers can be used to remove the gaps in the packets, but jitter buffers themselves cause additional packet transit delays in the network. When these delays become excessive, QoS becomes difficult to maintain.
Packet Loss and Out of Order Packets

IP networks cannot guarantee delivery of every packet, much less their order of arrival. Packets will be dropped under peak load conditions and during periods of congestion.

Among the approaches used to compensate for packet loss are interpolation of speech by replaying the last packet, and sending redundant information.

Out-of-order packets are treated as irrevocably lost and are replaced by their predecessors. When the late packet finally arrives, it is discarded. When packet loss rises above a certain level, QoS can no longer be assured.

Available Bandwidth

In a simple network consisting of only two end points, maximum transfer of data can be sustained and QoS can be guaranteed. In a larger real-world network, techniques are used to minimize congestion loss in the network, but these techniques reduce available bandwidth for an application. While plentiful capacity is a reasonable assumption for a controlled, localized environment, such as a corporate LAN, it is currently unrealistic across a global network such as the Internet. Higher levels of QoS require higher bandwidth requirements to support an application.

Service Level Agreements

Service Level Agreements (SLAs) are based on achievable levels of consistent QoS over a given network—meaning the ability of a given network to deliver the service needed by a specific network application from end-to-end. This can also include edge-to-edge, as in the case of a network that connects to other networks rather than to hosts or end systems, with the network itself providing some control over bandwidth, jitter, delay, and packet loss. Service Level Agreements apply most of the time.

Vendors promise SLAs with three levels of confidence: best effort, differentiated, and guaranteed.

1. **Best Effort Service.** Also known as lack of QoS, best-effort service is basic connectivity with no priorities or guarantees. It provides basic queuing during congestion with first-in, first-out (FIFO) packet delivery on the link. Examples of this type of traffic include a wide range of networked applications such as low-priority email and general file transfers.
2. **Differentiated Service.** Differentiated service means some traffic will be treated better (given higher priority) than the rest—faster handling, more bandwidth on average, and lower loss rate on average. Even so, high QoS is still not guaranteed.

Properly engineered, differentiated service can provide expedited handling appropriate for a wide class of applications, including lower delay for mission-critical interactive applications such as packet voice.\(^1\)

Typically, differentiated service is associated with packet classification, which means that traffic gets grouped or aggregated into a small number of classes, with each class receiving a particular QoS in the network. Thus, high-priority critical traffic receives a higher level of QoS assurance, while lower priority traffic receives a lower level. Differential service QoS is a reasonable trade-off between price and performance.

3. **Guaranteed Service.** Guaranteed service QoS requires an absolute reservation of network resources, typically bandwidth, which calls for reservation of buffer space, appropriate queuing disciplines, and so on, to ensure that specific traffic receives a specific service level. Guaranteed service is intended for delay-sensitive traffic, such as voice and video.\(^1\) Guaranteed service QoS costs the most.

### Improving QoS

Now that we've looked at the variables affecting QoS, let's explore some of the technologies that can be employed to improve network QoS for your enterprise. Several mechanisms can be used to support real-time and multimedia traffic at different layers of networking.

### Data Link Layer (Layer 2 of the OSI Model\(^2\))

At this layer, data packets are encoded and decoded into bits. The data link layer furnishes data protocol knowledge and management and handles errors in the physical layer, flow control, and frame synchronization. At Layer 2, media access control can be modified to provide service differentiation so QoS guarantees can be supported.

<table>
<thead>
<tr>
<th>OSI Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layer</strong></td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
Technologies employed to improve QoS at the data link layer include Asynchronous Transfer Mode (ATM), which is used in both wide area networks (WANs) and in local area networks (LANs); Frame Relays (FRs), which are used in the WANs; and IEEE 802 standards that are applied to the LAN media.

1. **ATM** is a network technology based on transmitting data in cells or packets of a fixed size. ATM can provide data-transport speeds in excess of 155 Mbps. As well as a high-speed bit-rate clock, ATM also provides a complex subset of traffic-management mechanisms, Virtual Circuit (VC) establishment controls, and various associated QoS parameters for these VCs. ATMs predominate in today’s Internet networks simply because of the high data-clocking rate and multiplexing flexibility available with ATM implementations.

2. **ATM Constant Bit Rate (CBR)** service is employed on virtual circuits that must transport traffic at a consistent bit rate, where there is an inherent reliance on time synchronization between the traffic source and destination. CBR service can be tailored for any type of data for which the end-systems require predictable response time and a static amount of bandwidth continuously available for the lifetime of the connection.

These applications include services such as video conferencing, telephony (voice services), or any type of on-demand service, such as interactive voice and audio. For telephony and native voice applications, AAL1 (ATM Adaptation Layer 1) and CBR service are best suited to provide low-latency traffic with predictable delivery characteristics.

3. **Frame Relay (FR)** is a link layer protocol that attempts to provide a simple mechanism for arbitration of network oversubscription. Frame Relay decouples the characteristics of the network access link from the characteristics of the virtual circuits that connect the access system to its group peers. Each virtual circuit is configured with a traffic Committed Information Rate (CIR), which conforms to a commitment on the part of the network to provide traffic delivery. However, any virtual circuit can also accept overflow traffic levels—bursts which may transmit up to the rate of the access link.
4. IEEE 802.1p provides a method to allow preferential queuing and access to media resources by traffic class, on the basis of a “priority” value signaled in the frame. This value will provide, across the subnetwork, a consistent method for Ethernet, token ring, or other MAC-layer media types. The priority field is defined as a 3-bit value, resulting in a range of values between 0 and 7, with 0 assigned as the lowest priority and 7 indicating the highest priority. Packets may then be queued based on their relative priority values.

5. IEEE 802.11 (Wireless LAN) specifies an over-the-air interface between a wireless client and a base station or between two wireless clients. This Wi-Fi technology is rapidly developing (802.11a, b, g, n, ac) to provide higher Mbps rates, greater bandwidth, and multiple-input multiple-output streams (channels), and greater assurance of end-to-end QoS.¹ ⁶

Over-provisioning

An alternative to complex QoS control mechanisms is to generously over-provision a network based on peak traffic load estimates. This approach is simple for networks with predictable peak loads. Performance is reasonable for applications such as video streaming, which can compensate for variations in bandwidth and delay with large receive buffers.⁷

Over-provisioning is of limited use in the face of transport protocols (such as TCP) that over time exponentially increase the amount of data placed on the network until all available bandwidth is consumed and packets are dropped. Such “greedy” protocols tend to increase latency and packet loss for all users.

Commercial VoIP services are often competitive with traditional telephone service in terms of call quality even though QoS mechanisms are usually not in use on the user’s connection to their ISP and the VoIP provider’s connection to a different ISP. Under high load conditions, however, VoIP may degrade to cell-phone quality or worse. The mathematics of packet traffic indicate that the network requires just 60% more raw capacity under conservative assumptions.

The amount of over-provisioning in interior links required to replace (or improve) QoS depends on the number of users and their traffic demands. This limits usability of over-provisioning. Newer, more bandwidth-intensive applications and the addition of more users results in the breakdown of over-provisioned networks. This requires a physical update of the relevant network links which is an expensive process. Thus over-provisioning cannot be blindly assumed on the Internet.⁷

By using QoS mechanisms, network administrators can employ existing resources efficiently and ensure the required level of service without reactively expanding or over-provisioning their networks.
Network administrators can use QoS to guarantee throughput for mission-critical applications so that their transactions can be processed in an acceptable amount of time. Network admins can also use QoS to manage User Data Protocol (UDP) traffic. Unlike Transmission Control Protocol (TCP), UDP is an inherently unreliable protocol that does not receive feedback from the network and, therefore, cannot detect network congestion.

Network administrators can use QoS to manage the priority of applications that rely on UDP, such as multimedia applications, so they have the required bandwidth even during periods of network congestion.

The Role of the Router

QoS depends on support throughout the entire network, end-to-end. To achieve QoS from sender to receiver, all of the network elements through which a traffic flow passes—such as network interface cards, switches, routers, and bridges—must support QoS. If a network device along this path does not support QoS, the traffic flow receives the standard first-come, first-served treatment on that network segment.

Routers supporting differentiated service configure their network scheduler to use multiple queues for packets awaiting transmission from bandwidth constrained (e.g., wide area) interfaces. Router vendors provide different capabilities for configuring this behavior, to include the number of queues supported, the relative priorities of queues, and bandwidth reserved for each queue.

In practice, when a packet must be forwarded from an interface with queuing, packets requiring low jitter (e.g., VoIP or video conferencing) are given priority over packets in other queues. Typically, some bandwidth is allocated by default to network control packets (such as Internet Control Message Protocol and routing protocols), while best-effort traffic might simply be given whatever bandwidth is left over.

Solutions

There are many reasons why achieving a reliable measure of network QoS is both essential and at the same time difficult. One compelling example of the need for QoS on the Internet relates to congestion collapse. The Internet relies on congestion avoidance protocols, as built into Transmission Control Protocol (TCP), to reduce traffic under conditions that would otherwise lead to “meltdown.”

QoS applications such as VoIP and IPTV, because they require largely constant bit rates and low latency, cannot use TCP and cannot otherwise reduce their traffic rate to help prevent congestion. QoS contracts limit traffic that can be offered over the Internet and thereby enforce traffic shaping that can help prevent overloading, and are hence an indispensable part of the Internet’s ability to handle a mix of real-time and non-real-time traffic without meltdown.
When selecting a vendor for your mission-critical over-the-Internet applications, investigate the vendor’s incorporation of technologies that improve (although do not guarantee) Quality of Service. In particular, the network router at your end is a critical component of QoS. Therefore, select a vendor that provides a list of recommended routers with a high level of reliability, capacity, and the ability to configure QoS settings.

A good practice is to test drive the vendor’s ability to deliver both capacity and quality of connection that you need. At least one vendor, RingCentral, provides online interactive testing of their VoIP service end-to-end, meaning from you-to-them-and-back, over your existing Internet connection and router. You may discover that VoIP will work reliably with your present connection and equipment, or that you would benefit from an upgrade.

Reliable call quality is an achievable goal when you, your ISP vendor, VoIP provider, and the greater Internet consistently employ Quality of Service disciplines and techniques end-to-end.

Industry-leading VoIP Technology from RingCentral

RingCentral delivers reliable, high-quality VoIP telephone service to hundreds of thousands of business customers. RingCentral has earned top industry honors, including the PC Magazine Editor’s Choice Award, and the 2010 World Economic Forum Technology Pioneer Award. Most recently, RingCentral won a CRM Excellence Award, and the 2014 Internet Telephony Product of the Year Award.

RingCentral delivers reliable business-class phone and fax services over the Web via its patented core technology infrastructure and global network housed in multiple state-of-the-art data centers.
Cited References

1 Quality of Service (QoS) in the New Public Network Architecture

2 The OSI Model
   http://www.firewall.cx/networking-topics/the-osi-model.html

3 Quality of Service on the Internet: Fact, Fiction, or Compromise?
   http://www.isoc.org/inet98/proceedings/6e/6e_1.htm

4 Introduction to ATM
   http://www.docstoc.com/docs/7491566/ATM-Introduction

5 Introduction to Frame Relay
   http://networklessons.com/frame-relay/introduction-to-frame-relay-for-ccna-students/

6 Workpackage 2 – Radio Communications

7 Quality of Service

8 What is QoS?

General References and Recommended Reading

QoS Architecture and Protocols for Internet Services

Understanding Jitter in Packet Voice Networks

The Steady Growth of VoIP Services
   http://briefmobile.com/steady-growth-voip-services

VoIP Meets Wi-Fi: Internet Phone Calls Go Wireless

Does Mobile Web Performance Optimization Still Matter?
   http://www.mobify.com/blog/web-performance-optimization/

Data Structures – Priority Queues
   http://web.augsburg.edu/~crockett/210/210%20Lab%20Set%2006/Lab0.html

Voice Over IP Basics
   http://www.voiptroubleshooter.com/basics/index.html

Measuring Voice Quality
   http://www.voiptroubleshooter.com/basics/mosr.html

Indepth: Jitter
   http://www.voiptroubleshooter.com/indepth/jittersources.html

Quality of Service

The 7 Layers of the OSI Model

Wireless Networking Standards
   http://www.webopedia.com/quick_ref/WLANStandards.asp

QoS for VoIP
   http://blog.simwood.com/2013/01/qos-for-voip-2/

Frame Relay Technology
   http://www.arcelect.com/frame_relay-56kbps_ft1-t1.htm
About RingCentral

Since 2003, **RingCentral** has been breaking down the communication barriers created by complex on-premise hardware. RingCentral’s cloud phone system delivers business communications solutions that free people to work the way they want in today’s mobile, distributed and always-on workforce. Delivered on a state-of-the-art cloud infrastructure, RingCentral helps more than 300,000 customers thrive in a new world of work. Learn why.