Voice over IP (VoIP) for Wireless Broadband Network Engineers

by Ken Rahmes

When discussing voice traffic across an IP based network, it is important to be able to quantify in some fashion the capacity required by a typical telephone conversation. Needless to say, this is a difficult proposition at best. Traditional methods deal with traffic intensity (in units of Erlangs or Centrum Call Seconds), but this paper will focus on the design of a data link based on the capacity required by a typical VoIP conversation. Figure 1 introduces some of the terminology used subsequently.

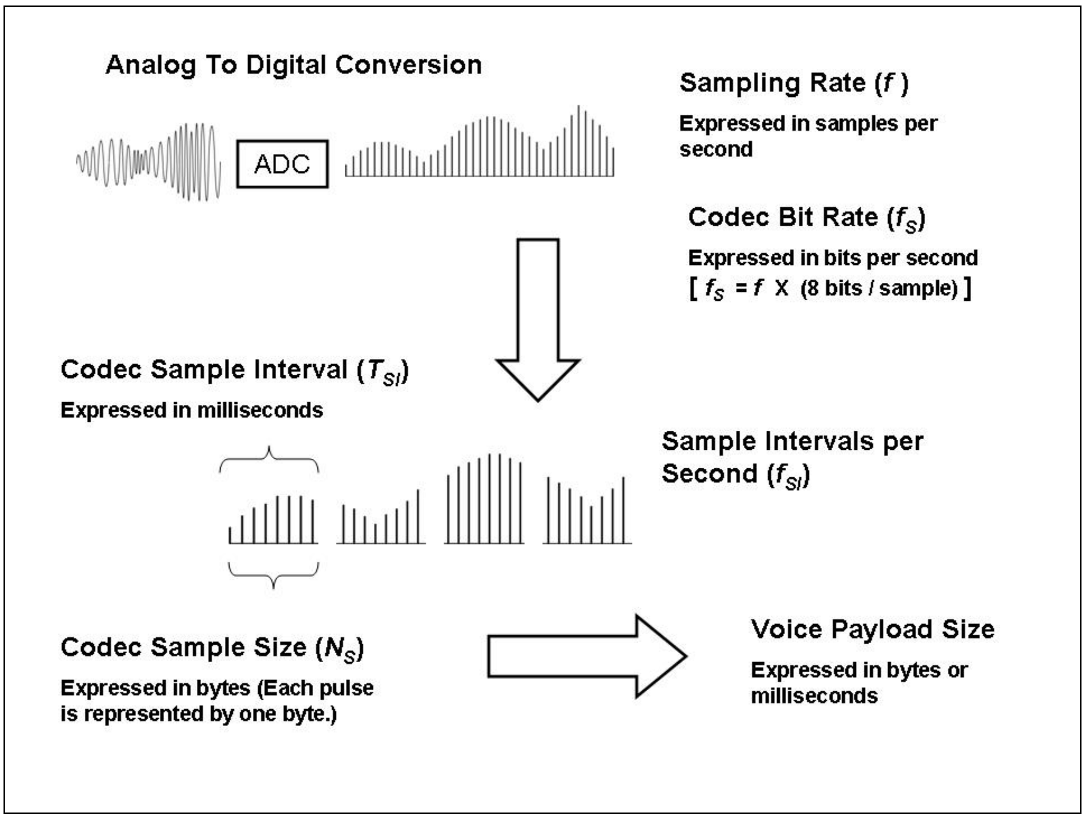

Figure 1: ***Codec Operation & Voice Payload***

From Figure 1, the following algebraic relationships are self-evident:

$$f_S = 8f \text{ } bps,$$

$$T_{SI} = \frac{N_S}{f} \text{ } ms,$$

$$f_{SI} = \frac{1}{T_{SI}} = \frac{f}{N_S} \text{ } SIps,$$

where,

f	sampling rate (Bps),
f_S	(codec) bit rate (bps),
N_S	(codec) sample size (Bytes),
T_{SI}	(codec) sample interval (ms),
f_{SI}	sample intervals per second (SIps).

The sampling rate (f) is actually measured in something like pulses or points per second, but the quantity is the same as the number of (binary) bytes, as each pulse is coded in one byte. So bytes per second (Bps) are used for simplicity herein.

High Quality & Low Data Rates

These days, everything traveling over the Internet faces the trade-off between maintaining original data quality and minimizing the amount of data transmitted. Audio data is no exception. Table 1 was developed using the foregoing equations starting with the codec bit rate (f_s) and sample interval (T_{SI}). The variable nature of the outputs of today's codecs is most obvious in the filling of Real-time Transport Protocol (RTP) packets.

For a given codec sampling rate and sample size, the insertion of one or multiple samples in a single RTP packet can vary the efficiency and quality of voice communication – unfortunately, inversely! Sample voice payloads are shown in the rightmost column of Table 1. (For the G.711 codec, the first entry shows a payload of two samples; the second, a payload of three samples.) The advantage of including multiple samples in a single payload is the reduction in overhead achieved. The disadvantage is the increase in latency and the increased impact of lost packets.

Since the purpose of this paper is to determine a rough estimate of the capacity required on a wireless data link in order to accommodate a single telephone call, the details of codec processing, network routing of packets, and jitter buffer depth will be left for another day. We will assume a common setting for the voice payload in an RTP packet and go on from there.

Table 1: ***Various Data Rates, Sample Intervals, & Payload Sizes***

Audio Codec	DSP Sampling Rate (f)	Codec Bit Rate (f_S)	Codec Sample Interval (T_{SI})	Codec Sample Size (N_S)	RTP Packet Voice Payload Size (B / ms)	Packets Per Second
G.711 (Pulse Code Modulation)	8 KBps	64 Kbps	10 ms	80 B	160 B / 20 ms	50
	8 KBps	64 Kbps	10 ms	80 B	240 B / 30 ms	33
G.729 (CPU resource intensive)	1 KBps	8 Kbps	10 ms	10 B	20 B / 20 ms	50
G.723.1 (6.4 Kbps)	800 Bps	6.4 Kbps	30 ms	24 B	24 B / 30 ms	33
G.723.1 (5.3 Kbps)	660 Bps	5.3 Kbps	30 ms	20 B	20 B / 30 ms	33

802.3 Encapsulation

The transmission of data across a packetized data link is a complex business. Symbols represent bits that are grouped into frames, packets, segments/datagrams, and finally data. Each of these units is typically referenced using the OSI Model and/or the TCP/IP Model. Table 2 shows both models, the terminology, and layers for reference. So how does a VoIP bit stream find its way across an IP network? It takes many layers of overhead.

Table 2. The OSI & TCP/IP Models.

OSI Designation	Layer	Data Unit	TCP/IP Designation	Layer	Examples
Application	7	Data	Application (Data, Voice, Video)	5	DHCP, DNS, FTP, HTTP, RTP, L2TP
Presentation	6	Data			
Session	5	Data			
Transport	4	Segment	Transport	4	TCP, UDP
Network (IP)	3	Packet	Network (IP)	3	IP, BGP, IPsec, ARP, RIP
Data Link (MAC)	2	Frame	Data Link (MAC)	2	802.11/16, Ethernet, PPP
Physical (PHY)	1	Bits / Symbols	Physical (PHY)	1	Bits, Symbols

As shown in Figure 2, the voice data is first incorporated into a Real-time Transport Protocol (RTP) packet at the TCP/IP Application layer. The RTP packet header contains such information as the format of the data, a sequence number, and timestamp. It is16 bytes in size, and in turn, is wrapped in a User Datagram Protocol (UDP) segment or datagram. The UDP header is 8 bytes long.

The principal advantage to using UDP datagrams is their simplicity and low overhead. UDP does not ensure delivery of a datagram. If a datagram is lost, oh well, too bad! The voice data is lost, but the end effect will probably be minimal; the point is: it's faster than TCP-like protocols that do ensure datagram delivery. Especially with voice data, the need for low latency outweighs the occasional lost syllable! Next, the IPv4 packet encapsulation (20 bytes) occurs. VLAN tags (4 bytes) may be a part of the mix; sometimes to separate the traffic of different users and sometimes to implement Quality of Service (QoS) classes. Finally, the MAC header and Cyclic Redundancy Check (CRC) checksum encapsulate everything (18 bytes).

Given the G.711 codec sample size of 80 bytes (with a codec sample interval of 10 milliseconds) and a "packet payload" of 160 bytes per RTP packet (Figure 2), about 50 packets per second (160 x 8 x 50 = 64Kbps) are required to transport the data across an Ethernet network. Given the sum of the overhead per packet of 66 bytes, then the overhead in bits is 66 x 8 = 528 bits/packet. So, (160 B/pkt x 8 bpB) + 528 = 1,808 bits/packet. So the grand total of bits per second to be transmitted over Ethernet II is:

1,808 bits/packet x 50 packets/second = 90,400 bps =. 90.4 Kbps.

This is a peak rate, of course, after all the protocol setup and handshaking has taken place. It also does not take into account the physical layer (Layer 1) which will add a preamble and start frame delimiter (SFD) totaling 8 bytes.

There are other codecs (G.729, G.723.1, G.726, G.728) that produce lower data rates, but as network designers do not typically get to specify the codec used by clients, G.711 is a conservative choice when estimating necessary bandwidth.

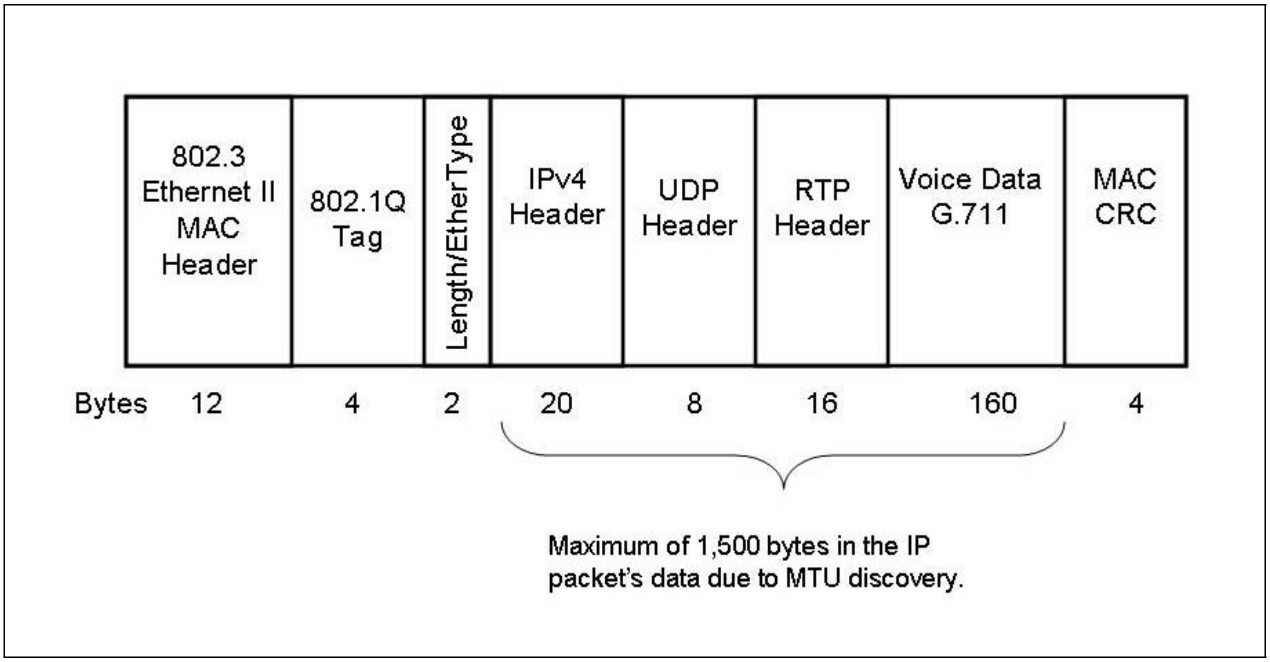

Figure 2: ***Ethernet VoIP Data Encapsulation***

802.11 Encapsulation

Data is fed to a wireless access point (AP) at the rate of about 90.4 Kbps in this scenario. What does the AP do with that data and what is the data rate required when transmitting through the mesh network?

First, any router, and an AP can be thought of as a wireless router, will strip off the MAC header, either remove, or more likely pass through the 802.1Q tag, and resolve the IPv4 address with the addresses in its routing table. The destination IP address will be across the mesh network on the other side of a gateway radio or POP. While there are many possibilities, one is that nothing changes in the payload. The only change occurs at TCP/IP Layer 2, the MAC frame.

Now, instead of an Ethernet II MAC, an 802.11a MAC frame header is prepended to the data and the 802.3 MAC CRC is replaced by an 802.11a Frame Check Sequence (FCS). The FCS is essentially the same as the 802.3 CRC checksum, but it's been modified to incorporate the changes made to the MAC header. The only real change from the perspective of data rates is that an 802.11 MAC header is 24-30 bytes while the 802.3 MAC header is 12 bytes. Also, 802.11 includes LLC/SNAP fields that add 8 bytes. So for each frame (f) 26 bytes (B) need to be added: 8 bpB x 26 Bpf x 50 fps = 10,400 bps. So, if 10.4 Kbps are added to 90.4 Kbps, that gives a rough bit rate for a mesh network link of 100.8 Kbps. This is not a huge difference! (Just as with the Ethernet physical layer, 802.11 requires some housekeeping. In this case it's a PLCP header and a preamble. But, when a radio manufacturer specifies a particular data rate, these PMD level issues are not included so they are not relevant to the issue of network capacity.)

In addition to the Frame Check Sequence (FCS), noted above, 802.11a adds a certain amount of error handling capability at the physical level, Layer 1 of the OSI and TCP/IP models. It is commonly referred to as Forward Error Correction (FEC). Typical ratios are 1/2, 2/3, and 3/4, where the numerator is the number of data bits transmitted and the denominator is the total number of bits including the FEC. So, for every three data bits one additional FEC bit is included when a 3/4 FEC rate is specified.

Now, at maximum capacity – 64-QAM with 3/4 FEC – a single OFDM subchannel moves 1.125 Mbps of data. Each link has 48 data subchannels, so that's 54 Mbps. Dividing by 100.8 Kbps, the estimate of the voice data rate for the G.711 codec determined above, results in about 535 simultaneous VoIP telephone calls!

Does this make sense? Is this reality? Actually, no. It turns out that in the real world, what with the error rates actually encountered – bit errors, dropped packets, etc. – an optimistic throughput is something less than half that 54 Mbps, say 20 Mbps! So, we're really talking something on the order of 200 simultaneous calls. Keep in mind, these are estimated "maximum" rates across a point-to-point data link! When you send VoIP data across a mesh network, data can degrade significantly, so beware, and as always: test, test, test.

An AP-to-AP wireless MAC frame requires four IP addresses, as shown in Figure 3. The first and second addresses are those of the receiver and transmitter on the wireless network. Then, the destination and source IP addresses follow as shown. MAC data frames in the 802.11 standard can accommodate up to 2312 bytes of data; considerably more than the 1500 bytes of the 802.3/Ethernet II standard. Unfortunately, the wireless network cannot concatenate multiple packets! So, the amount of voice data per frame remains the same.

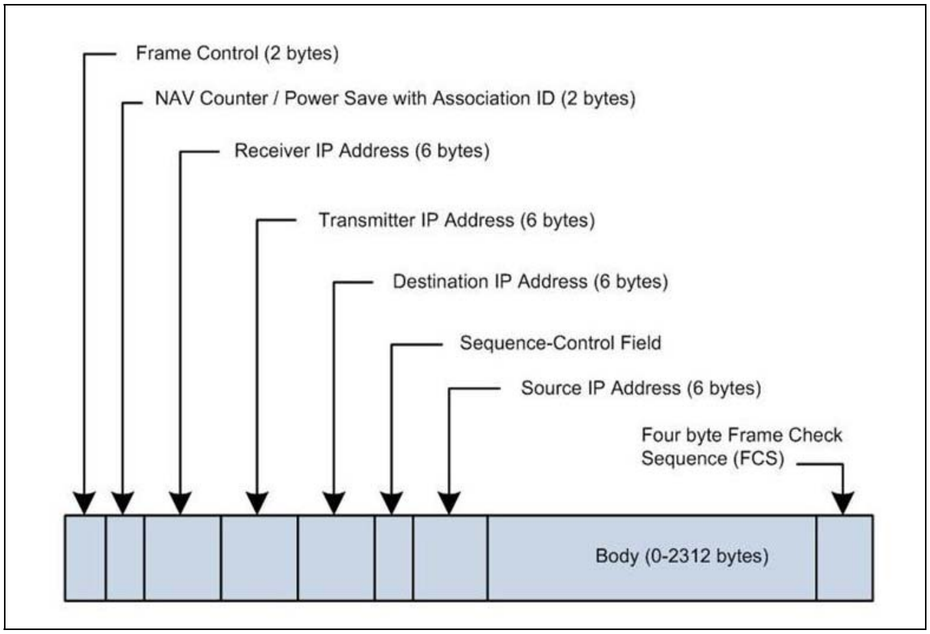

Figure 3: ***802.11 Data Frame - AP to AP***