Video Calling Over Wireless Networks

David Urban
Comcast

Abstract

Video calling over wireless networks has become increasingly popular as wireless networks become faster and more reliable and devices with cameras and video calling applications become more ubiquitous.

Measurement and analysis of video calls over home, outdoor and cellular wireless networks have determined the criteria for making a successful call over a wireless network. Signal strength, packet loss, jitter and round trip delay are critical parameters. Video calling over wireless networks is shown to be practical, provided that the critical parameters are met.

INTRODUCTION

Video calling is a technology that has been around for quite some time but has never caught on as much as one might think. As the saying goes, video calling is the technology of the future and always will be.

Bell Labs began building experimental prototypes in 1956 culminating in the 1964 New York World's Fair demonstration of the Picturephone service [1]. By 1969, the transition from voice calling to video calling appeared to be at the threshold. Looking back at the Picturephone service of the 1970s, it is interesting and instructive to find that many of the standards and specifications are similar to those used today. The analog bandwidth of the black and white video picture was 1 MHz with an interlaced 250 lines refreshed at 30 frames per second. The screen size was 5.5 x 5". When digitized to be transported a distance greater than 6 miles the combined video and audio signal was 6.3 Mbps.

So have things changed appreciably enough to suspect that this might just be the time that video calling really catches on? There is ample reason to think that the answer is yes. Many factors have fallen into place to make video calling more feasible than ever before.

Many people now carry around smart phones with a front and back camera, video calling software and a data connection fast enough for video calling. This means that when initiating a video call, one needn't count on the other side being at their computer with attached web camera and logged into a video calling application.

Televisions can be made into high definition video conferencing solutions with convenient and inexpensive add on products such as video cameras with built in microphones and small computer appliances to run the video calling application. Again, this avoids the inconvenience of having to fire up your computer, plug in your webcam, and open and log in to your video conferencing software. Any time you are watching television you can make or receive a video call.

A video call on a large screen television set can be much more enjoyable than using a computer. Several members of the family can participate while sitting on the couch rather than crouching around a small computer screen. And the 720P resolution of a 32 inch or larger diagonal flat screen television provides a much better viewing experience than a notebook computer or smart phone screen can.

Successful video calling requires a network connection with high data rate, low latency and jitter, and negligible packet loss.

Broadband connections are becoming more common and the performance keeps improving, making video calling more practical. This is true for both fixed and mobile networks. Video codecs have and continue to improve and work is being done specifically for video grade wireless distribution.

While there are still some impediments to video calling such as high cost and the lack of simple, intuitive and reliable user interfaces, many hurdles to successful video calling have recently been cleared and the remaining obstacles are trending toward resolution. In the 1960s and 1970s video calling moved from a laboratory curiosity to an ambitious but ultimately disappointing large scale national project. Then in the 1980s and 1990s video conferencing remained a niche application mainly for big businesses. With the advent of the personal computer and broadband residential network connectivity video calling has become an increasingly popular method to stay in touch with family and friends. The big transition occurring today is the move from video calling on desktop and notebook computers to video calling on smart phones, tablets and televisions. This transition makes wireless home and public network performance and reliability more important than ever. Table 1 shows some video call data rates.

Video Call Type	Tx Mbps	Rx Mbps	Video Quality
1080P WiFi	5	5	excellent
720P WiFi	1.5	1.5	excellent
3-Way WiFi	1	1	good
Smart Phone	0.5	0.5	fair
3G Cellular	0.2	0.2	poor

Table 1. Typical Data Rates of video calls

VIDEO CALLING PARAMETERS

A video call can be at times amazing when it works perfectly while at other times the experience can be frustrating when things go wrong. The elements of a video call include two parties, each with a video camera, video display, audio microphone, and audio speaker. Each party needs a computing device to run a video calling application and the devices need to have a network connection to establish the call, send the video and audio streams, and end the call.

For a two-way video call, a video stream will be sent from the video camera and another video stream will be received for the video display. Likewise, an audio stream will be sent from the microphone and another audio stream will be received by the speaker.

The audio and the video must be synchronized. A delay from the video camera of one user to the video display of the other user can be distracting. For example, if a caller wishes to show an object by putting it in front of the camera but gets no reaction from the other side, this can be confusing. Then the other side finally comments but long after the object has been removed from the camera view. This distracts from the real time interactivity of the video call.

Disconnects, long reconnections, poor video quality, long delays, lack of video and audio synchronization, freezing of the video, brief distortions of the video display, screen refresh and resolution issues; these are the problems that make video calling frustrating. A successful video call requires a good network connection on both ends, good processing power in the CPU running the application, a good video calling application, a good camera and display on both ends.

The key network connection parameters necessary for a successful video call include data rate, packet loss, jitter, delay, and relays. The data rate will be dependent upon the screen size and resolution. A video call on a 1080P LCD television will have a data rate of 10 Mbps whereas a video call on a 4.3 inch diagonal screen smart phone will have a data rate of 1 Mbps.

Video calling applications often report call technical information. Among the reported parameters are jitter, packet loss, send packet loss, receive packet loss, round trip time, and relays. Relays can be used to work around firewalls and other networking issues that prevent a direct UDP connection between the two video callers. Relays in general are undesirable since they often prevent HD video calling. The video and audio streams are sent as UDP packets.

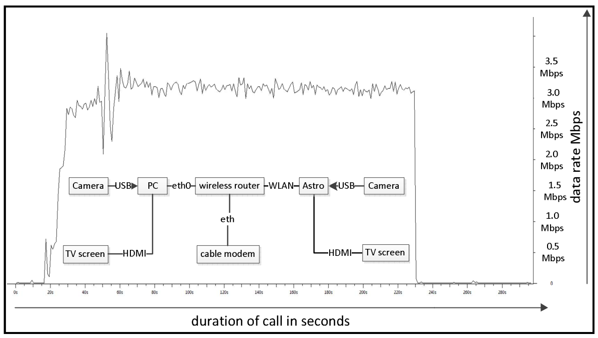

Fig.1 Data Rate and Block Diagram of 720P Video Call

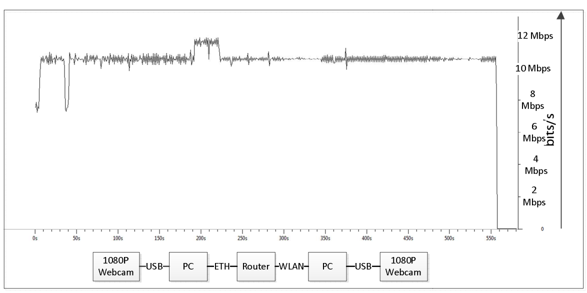

Fig. 2 1080P video call data rate

Wireshark was used to record the packets during a video call. The data rate during the call is shown in Fig.1 along with a block diagram of the test set up. Both the camera and the display were capable of 720P operation. The upstream and downstream data rate was measured to be 1.5 Mbps for a total of data rate of 3 Mbps. The video call quality was excellent. Packet analysis shows that the data protocol was UDP with packet size around 1400 bytes. In this particular test the video and audio streams were sent between devices on the same local area network. Most video calls will span a wide area network adding additional challenges for a successful video call.

Figure 2 shows a video call with 1080P video resolution. In this case the data rate is much higher at 10 Mbps, 5 Mbps for each video stream. Setting up a 1080P video call can be a bit tricky. You'll need a 1080P video camera and display at both ends, video calling software the supports 1080P resolution at both ends, and network connectivity supporting 5 Mbps UDP traffic in both the upstream and downstream direction. Residential broadband connections that support this high upstream data rate have only recently been offered. Figure 3 shows a speed test for a cable modem connection capable of supporting 1080P video calling.

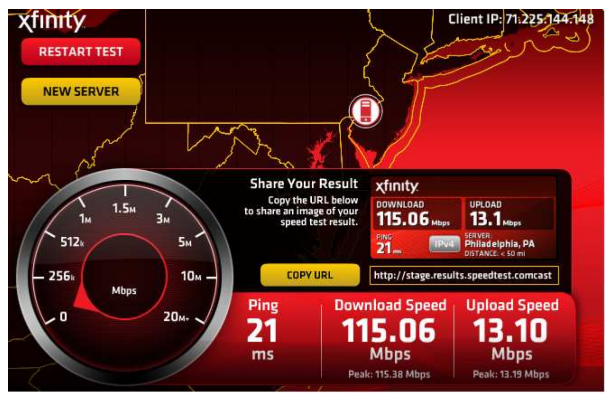

Fig.3 Broadband Connection speed for 1080P video call.

The user experience of a 1080P video call is remarkable. The picture is clear and sharp on a big screen television and the live fast action response to motion is impressive.

Large screens with high resolution benefit from very high continuous data rates during a video call; however, many video calls involve smart phones which have much smaller screens that do not need such high data rates. Figure 4 show the data rate measured during a video call using a smart phone. The smart phone network connectivity is over a wireless home network.

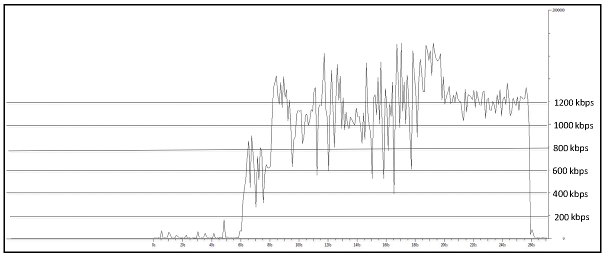

Fig. 4 Video Call using a smart phone with WiFi

The data rate measured about 1.2 Mbps and the quality of the video was good. Notice that the data rate is much less consistent than the previous plots with the bit rate over time being very choppy. This is due to several factors including the wireless home network link, the smart phone CPU processing speed and memory, and the impact of the wide area network. For the video call in figure 4 two cable modems were used so that the video and audio streams had to traverse the HFC network.

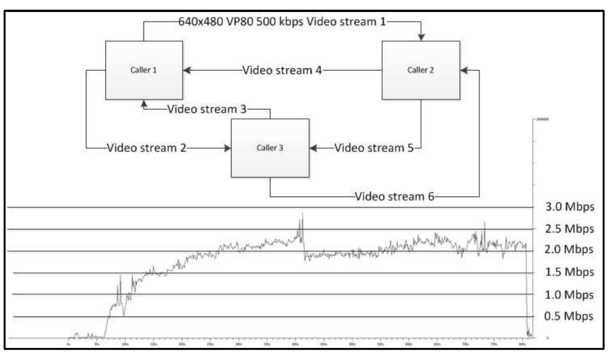

Fig. 5 Three-Way Video Call

A video call can be made between three or more parties. For a three way video call, a video caller sends two video streams and receives two video streams. The video callers' display typically shows the two received video streams side by side on the screen with a small caption of the video send stream. With this format the video caller can see both of the people he is calling as large as possible and still monitor what the other parties see of him. Since two video streams must share the display, the resolution and bit rate of a single video stream is reduced, i.e. one cannot display two 1080P video streams on a single 1080P video display. Testing 3-way video calls, the video send and receive streams were found to be 640x480 with VP80 codec at 30 frames per second and a bit rate around 500 kbps. A video caller participating in a 3-way call will thus send two 500 kbps video streams and receive two 500 kbps video streams for a total data rate of 2 Mbps as shown in figure 5.

VIDEO CALLING OVER WIRELESS NETWORKS

Characteristics of the wireless network

The making of a successful video call requires network connectivity with low packet loss, low latency, and low jitter and must support UDP data rates between 1 and 10 Mbps depending on the screen size and video quality requirements. Several wireless networks were tested to gauge their performance against the demands of video calling.

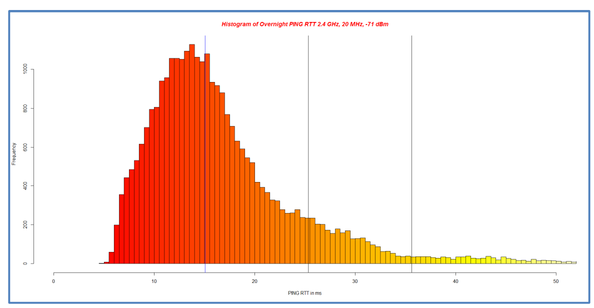

Fig. 6. Histogram of Overnight PING RTT 2.4 GHz, 20 MHz, -71 dBm from 0 to 50 ms

Fig. 6 shows a histogram of the round trip time, RTT, measured while sending PING packets between a wireless client and a wireless access point of a home wireless local area network, WLAN. The x-axis is the PING RTT from 0 to 50 ms and the y-axis is the number of occurrences. As indicated by the vertical lines in figure 6, the median RTT was found to be 15 ms, the first standard deviation above the median was 25 ms and the second standard deviation above the median was 35 ms.

The wireless access point was set to 2.4 GHz channel 8 with a 20 MHz channel width. Both the STA and the AP were IEEE 802.11n with dual stream capability. The wireless access point was set to B/G/N Mixed wireless mode. The beacon interval was set to 100. The RTS threshold was set to 2347. The guard interval was set to 800 ns. The STA and AP were separated by 36 feet and one floor and two walls of a residential home. The PING tests were taken over a 12 hour period. The x-axis of the plot is the PING round trip time measured in ms. There are three vertical lines shown on the graph, from left to right these lines are the median, first standard deviation, and second standard deviation, respectively. The receive level measured by the STA was -71 dBm.

The results indicate that the latency and jitter of a wireless home network have much more variability than a wired network over the course of time. This can be due to signal fading and interfering signal sources. The statistical distribution of the round trip time of packets between the AP and the STA are well within the requirements of a video call. A round trip time of 40 ms will support a high quality video call. On the histogram of round trip time measured in ms, 40 ms is beyond the second standard deviation and thus is a rare occurrence.

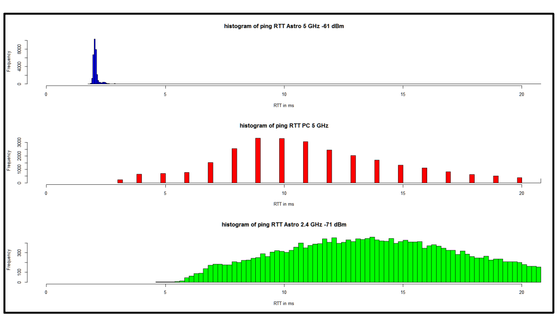

Fig. 7 Histogram of PING RTT for 5 GHz -61 dBm(top), 5 GHz at same location with PC (middle), and 2.4 GHz -71 dBm (bottom) from 0 to 20 ms

Figure 7 shows test results of the distribution of PING round trip time in ms over the course of a 12 hour test period. There are three different test conditions, the top blue graph is the PING RTT distribution over a 12 hour test period of a 5 GHz wireless home network connection with a receive level of -61 dBm. The computer used for this test was a small form factor LINUX device with built in WiFi client. For this test the AP and the STA were separated by one wall and 12 feet. 5 GHz band with the AP and STA in close proximity results in a much lower median round trip time latency of 2 ms with no significant measurements greater than 5 ms.

The middle red distribution of figure 7 shows another 5 GHz test taken in the same location and same time as the top distribution but using a different wireless client station. The middle test results used a notebook computer with built in wireless card and antennas. This RTT distribution shows a median of 10 ms with most RTT measurements fewer than 20 ms.

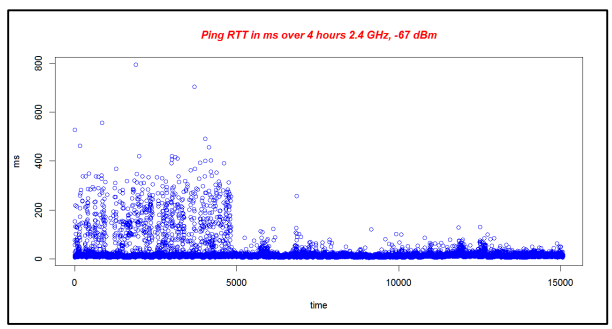

Fig. 8 Plot of PING RTT in ms over 4 hours at 2.4 GHz, -67 dBm, y-axis is RTT in ms from 0 to 800ms.

Both test results are good and well within the requirements for a successful video call. But why would two tests, both wireless clients using the same channel at the same time, both in the same location, give such different results? It turns out that it was not due to hardware differences between the two stations since subsequent overnight tests revealed that by slight manipulations of the antenna positioning one could reverse the results.

As illustrated in figure 8 the PING round trip time can change abruptly in time and these changes can last for hours at a time. This could be due to other applications sharing the spectrum or even to the movement of people or objects within the home.

The antenna patterns of notebook computers and small form factor devices with built in antennas will have nulls due to internal obstructions. By slightly repositioning the computers and devices one can place the nulls in more or less advantageous a location and this can influence the PING RTT results.

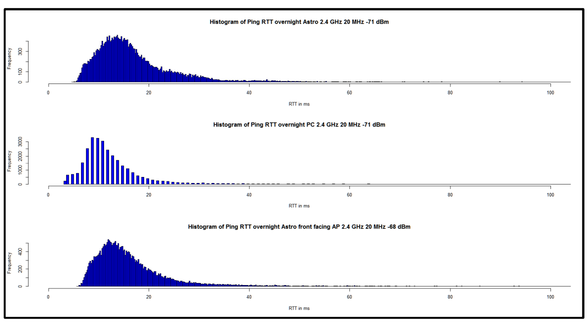

Fig. 9 Histogram of PING RTT overnight at (top) 2.4 GHz, 20 MHz, -71 dBm, LINUX, (middle) 2.4 GHz, 20 MHz, -71 dBm, WINDOWS, (bottom) 2.4 GHz, 20 MHz, -68 dBm, LINUX with antenna facing AP, x-axes are RTT in ms from 0 to 100.

The bottom green PING RTT overnight test distribution in figure 7 was taken using 2.4 GHz with the STA receive level of -71 dBm due to a larger separation distance between AP and STA of 36 feet with one floor and two walls. As expected the PING RTT distribution is much larger than the test using 5 GHz at closer AP to STA separation distance with a median of 14 ms and a significant number of round trip times having latency greater than 20 ms. The 2.4 GHz and -71 dBm receive level overnight PING test shows performance that is well within the bounds for a successful video call but as we will later see this is at the threshold of successful video calling operation.

One can expect variety of latency and jitter distributions for wireless home networks. Other examples are shown in figures 9 and 10. Figure 9 shows three tests taken in the same location, all at 2.4 GHz with 20 MHz channel width. The difference between the three plots is due to slight differences in antenna positioning. Figure 10 shows a comparison of 2.4 GHz performance versus 5 GHz at a 24 foot AP to STA distance with one wall in between. At close range 5 GHz proved to have consistently lower round trip times, however, figure 10 shows that at farther distances and more wall attenuation 2.4 GHz operation can have lower round trip time than 5 GHz.

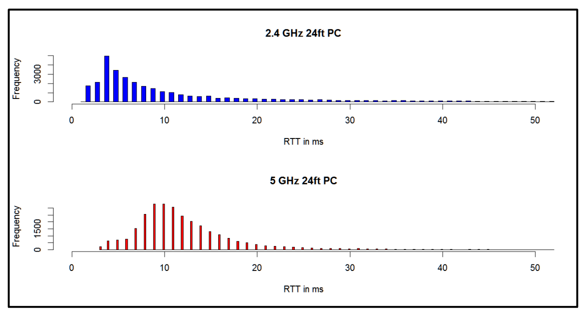

Fig. 10 Histogram of PING RTT 2.4 GHz vs 5 GHz and 24 ft AP to STA distance

In general wireless connections will be worse than wired connections in this regard. It is important to note that some routers handle IP video and peer to peer stream video better than others. As a rule of thumb, using 5 GHz at very close distance will give more consistent performance for video calling than 2.4 GHz at far separation distances as illustrated in figure 7. If you are having trouble making a video call using a wireless home network connection, then slight variations in antenna positioning of either the client station or access point can make significant performance improvements. Changing the RF channel, channel width, guard interval, and mode may also be experimented with to fix problems.

Figure 11 shows the statistical distribution of the call technical information reported by the video calling application. A small form factor Linux computer was used to run the video call application. The video camera and display at both ends of the video call were 720P and the video call data rate was 3 Mbps. The wireless network connection used 2.4 GHz with a 36 feet AP to STA separation distance with one floor and two walls in between. The video calling software has an option to report call quality technical information which includes a measure of network packet loss, roundtrip time, and jitter. These statistics are used to adjust the call quality to account for network connectivity issues. If these parameters degrade, then the video calling application will adjust by lowering the video quality such as adjusting the resolution from 1280x720 to 640x480. Adaptive bit rate streaming is a technique to provide the best video quality for given network limitations.

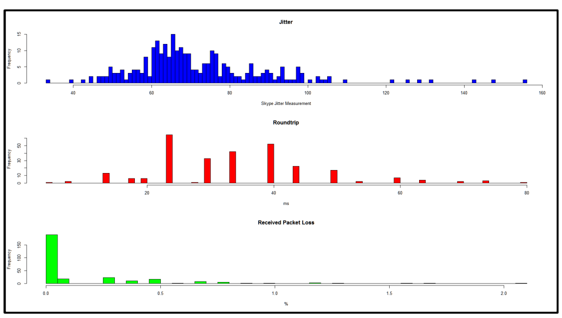

Fig.11 Video Call Quality Technical Information Top Histogram is Jitter from 40 to 160, Middle Histogram is Roundtrip time from 0 to 80 ms, Bottom Histogram is received packet loss from 0 to 2%

The call quality technical information was saved to a text file during the video call. A PERL language program was written to filter out the three parameters of interest into an array suitable for statistical analysis using the R statistical programming language. The analysis shows that the packet loss throughout the video call remained low at less than 0.5%. The round trip time varied significantly with a noticeable amount of measurements as high as 60 ms. The jitter measurement also varied significantly during the course of the call. Despite the variations, test results show that the wireless network was able to support a 720P video call with good reliability.

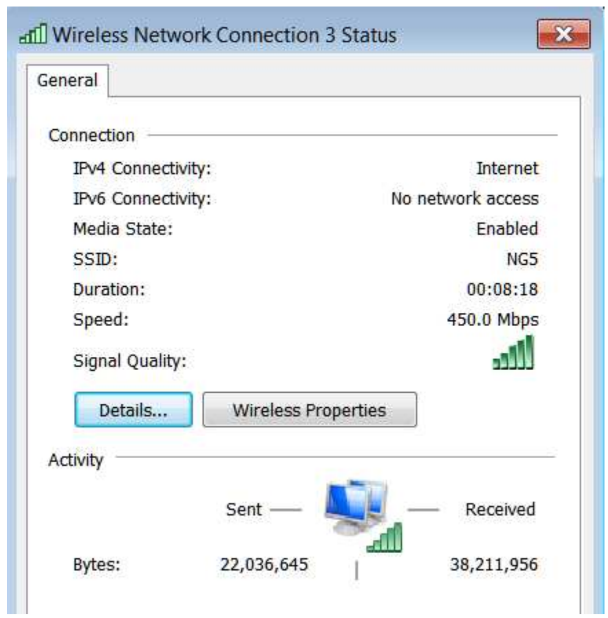

Fig. 12 3by3 AP and 3by3 STA reporting 450 Mbps modulation and coding scheme

Wi-Fi packet analysis of a 1080P video call

A video call was set up between two callers with one caller using a wireless home network. Both the access point and the client station of the wireless home network were capable of three stream operation. The highest data rate of the wireless home network was 450 Mbps. By carefully positioning the access point and the client station in close proximity and applying some tricks such as using cookie sheets to create reflections it was possible to get the client wireless software to report 450 Mbps as shown in figure 12.

However, during the video call the highest data rate achieved was 324 Mbps. The data rate of 324 Mbps has three spatial streams, a guard interval of 800 ns, a 40 MHz channel width, and MCS 21 64-QAM with 2/3 rate binary convolutional coding. The method to calculate the data rate of 324 Mbps is shown in equation 1. The details behind these calculations can be found in [2], [3],[4].

$$R = \frac{3(streams) * 6\left(\frac{bits}{sc}\right) * \left(\frac{2}{3}\right) * 108(sc)}{(3.2 + .8)(\mu s)} = 324\ Mbps\ [1]$$

With three transmit and three receive antennas in the access point and the client station there are nine paths between the transmit antennas and the receive antennas as shown in figure 13.

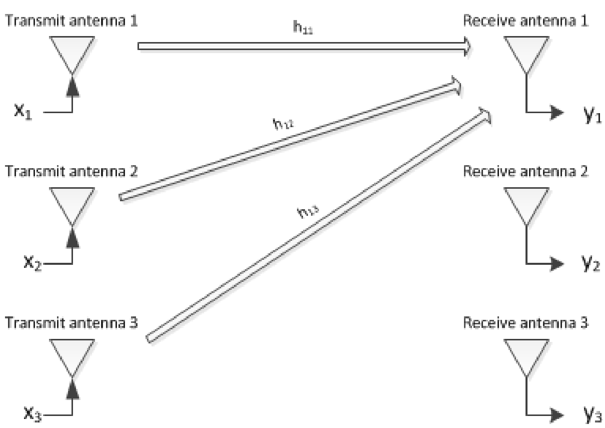

Fig 13. 3x3 MIMO Block Diagram

The output signals of the receive antennas, y_i with i={1,2,3}, is equal to the input signals of the transmit antennas, x_i, times the complex path loss of the nine paths between transmit and receive antennas, h_{ij}, as shown in figure 13. The relationship between the output signal of the three receive antennas and the path loss between the antennas and the input signals to the three transmit antennas can be expressed as a matrix equation [2]. If the H matrix of equation [2] can be inverted then it is possible to calculate the input signals by measuring the output signals and multiplying by the inverse of the H matrix. The determinant of the H matrix is zero if all of the elements are the same. The inverse of the H matrix is proportional to the inverse of the determinant. Thus, if all the elements of the H matrix are identical the determinant will go to zero and the inverse will blow up to infinity and it will not be possible to determine the input signals with knowledge of the output signals and path characteristics. Multiple streams can only work if there are differences, most desirably phase differences, between all of the nine paths between antennas. Spreading the antennas apart spatially is one method to increase the phase differences between the paths. However, with compact access points and particularly with compact client stations the amount of spatial separation is limited. Here, 5 GHz operation has an advantage over 2.4 GHz operation since for a given spatial separation, electrically in terms of wavelengths the separation between antennas is greater at 5 GHz than 2.4 GHz. The most effective and desirable method to create differences between the paths is reflections. A multipath rich environment with many reflected signals is the best for realizing multiple streams of data. In equation [1] the data rate from a signal antenna is multiplied by 3 since each of the 3 transmit antennas are transmitting an independent data stream.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\ [2]$$

Each subcarrier of the OFDM symbol is 64-QAM modulated so that a subcarrier is mapped to 6 bits. The bits are the output of a binary convolutional coder that inputs 2 data bits and outputs 3 coded bits. Thus, each OFDM subcarrier is mapped to 4 data bits as reflected in equation [1].

The channel width observed for this test video call for this packet was 40 MHz. A 40 MHz 802.11n signal consists of 128 subcarriers. Subcarriers at the channel edges and center are nulled to form a guard band and prevent DC offset. Some subcarriers are used as pilots to allow for frequency acquisition and carrier lock. This leaves us with 108 data subcarriers for a high throughput 802.11n data packet as reflected in equation [1]. Equation [1] reveals that each OFDM symbol for the observed packet has 1,296 data bits or 162 data bytes.

Finally, in order to determine the data rate we need to know the symbol time. The 128 modulated subcarriers that comprise the OFDM symbol are converted to a time domain representation using an inverse fast Fourier transform, IFFT. The 128 point IFFT is a transform with 128 complex frequency domain input numbers and 128 complex time domain output numbers. A digital to analog conversion, ADC, is required to turn the complex numbers into a real waveform capable of being upconverted to a carrier frequency to excite an antenna current in order to form an electromagnetic wave that can radiate from the transmit antenna to the receive antenna. The clock of the ADC determines the channel width of the analog time domain waveform. The channel width will be 40 MHz if the sampling interval is 25 ns. The formula is shown in equation [3] with W being the channel width in Hz and τ is the sampling interval of the time domain waveform in seconds.

$$W = \frac{1}{\tau} \quad [3]$$

With 128 subcarriers turned into 128 time domain samples by an IFFT and applying a 25 ns sampling interval, an OFDM symbol can be transmitted in 3.2 μs. This is sometimes referred to as the useful symbol rate. In theory OFDM symbols could be sent every 3.2 μs, however, in practice a guard interval in the form of a cyclic prefix is added so that the OFDM symbols are sent at an interval that is longer than the minimum possible. This eliminates inter-symbol interference that results when the receiver is hit with two different symbols at the same time due to reflections. As long as the guard time is longer than the time delay of the largest reflection then the receiver can ignore the guard time and demodulate the useful symbol time without inter-symbol interference. For 802.11n OFDM symbols the guard interval, GI, can be either 800 ns or 400 ns. A GI of 400 ns is referred to as a “short guard interval.” For the packet analyzed in the example the guard interval was 800 ns or 0.8 μs. The total OFDM symbol time is the sum of 3.2 μs and 0.8 μs for a total symbol time of 4 μs. This is shown in equation [1]. Now the data rate can be calculated by dividing the bits per OFDM symbol by the symbol time, in this case the data rate is 324 Mbps.

The frame length of the QoS data packet was 1395 bytes so that 9 OFDM symbols carrying 162 bytes each are needed to send the packet data payload. A burst of 9 OFDM HT symbols in this example lasts 36 μs since the OFDM symbol time for normal guard interval is 4 μs. Thus, over the 36 μs period of the 9 OFDM HT symbols the data rate is 324 Mbps.

When digital video signals are sent from a cable headend to a receiving set top box, the 256-QAM modulated 6 MHz wide signal transmits 38 Mbps continuously, a 100% duty cycle. This is not the case for wireless local area network transmissions. Since the medium is shared between uplink and downlink transmissions, amongst other users of the wireless home network, amongst co-existing wireless home networks, and amongst other spectrum users such as microwave ovens, cordless phones, remote controls, and sensors, a 100% duty cycle is not possible. Data is sent in bursts with short time frames and these bursts require a preamble in order to be received. The preamble is needed in order for the receiver to acquire carrier lock, understand the basic parameters of the packet, so that demodulation of the payload symbols can be made accurately.

All packets must be preceded with a preamble. The first part of the preamble is a short training field made up of 12 subcarriers. The short training field is 8 µs long. The short training field consists of 12 subcarriers. Figure 15 shows the subcarriers of the short training field measured by a vector signal analyzer. The short training field is followed by and 8 µs long training field and then a 4 µs signal field.

The transmission of a video packet of from the AP to the STA requires a sequence of packets as shown in figures 14 and 15. First, the AP sends a request to send message to the STA. The STA responds with a clear to send message. A QoS Data packet is sent from the AP to the STA. Finally, a block acknowledgement message is sent from the STA to the AP. This process is repeated continuously throughout the video call for both uplink and downlink transmission.

The request to send packet reported a length of 16 bytes and a data rate of 24 Mbps in its signal field, labeled SIG in Fig. 14. This is a legacy packet and thus has a 20 µs preamble consisting of an 8 µs short training field, STF, an 8 µs long training field, LTF, and a 4 µs signal field. The symbol period is 4 µs, a symbol has 48 data subcarriers mapped to 2 data bits so each symbol carries 96 bits or 12 bytes of data. The RTS packet is 28 µs and the request is to transmit a packet sequence with 224 µs duration.

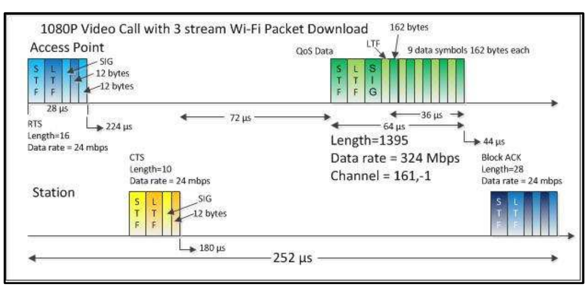

Fig. 14 WiFi Downlink video packet sequence

Following the RTS from the access point to the station will be a clear to send, CTS, response from the station to the access point. The clear to send packet has a length of 10 bytes and a data rate of 24 Mbps with a channel of 161. As with the RTS, the CTS packet has an 8 µs short training field, followed by an 8 µs long training field, followed by a 4 µs signal field, followed by 4 µs OFDM data symbols carrying 12 bytes of data. Since the CTS field length is 10 bytes only one OFDM data symbol is needed.

The CTS packet time duration is 24 µs. The CTS signal field reports that the duration from the end of the CTS to the end of the packet sequence is 180 µs. By taking the difference between the reported duration by the RTS packet and the CTS packet, we calculate the time duration from the end of the RTS packet to the end of the CTS packet of 44 µs. Thus there is a gap of 12 µs from the end of the RTS packet to the beginning of the CTS packet allowing for time for the access point request to be made and the client station response to be sent.

After the access point makes a request to transmit data and the client station responds with a clear to send signal then the QoS data packet can be sent from the access point to the client. Once the QoS packet has been sent by the access point and received by the client station then the client station sends a block acknowledgement back to the access point.

So in this example packet sequence measured during a 1080P video call, 1395 data bytes were transmitted over a 252 µs time period. The data rate accounting for the signaling and overhead is 1395 bytes divided by 252 µs which is 44 Mbps. Since the 1080P video call requires a sustained 10 Mbps data rate, the duty cycle of 324 Mbps data rate QoS data packet sequences during a 1080P video call is 23%.

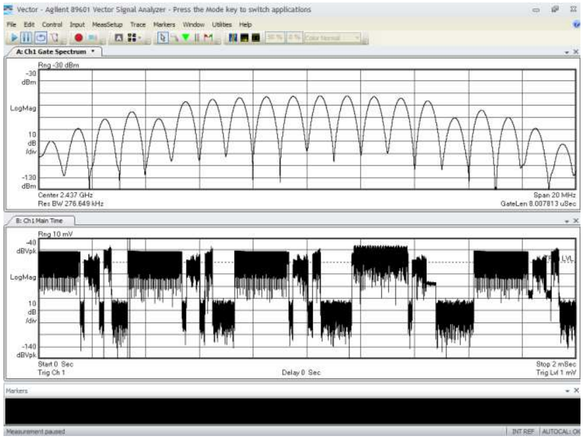

Fig. 15 Spectrum Analysis of WiFi video call

Taking a look at the distribution of the data rate of QoS Data packets reveals that many of the data packets were sent at a lower data rate than 324 Mbps. During the entire 1080P video call 3 stream operation was only utilized a small percentage of the time. All in all, 392,129 packets were analyzed. Of all of the downlink QoS data packets 9.42% had a data rate of 324 Mbps utilizing 3 stream operation. The majority of downlink QoS data packets operated at a 2 stream data rate of 243 Mbps representing 74.92% of the downlink QoS data frames. On the uplink 85.59% of the QoS data frames had 2 stream 270 Mbps while only 0.4% of uplink packets used 3 streams at a data rate of 324 Mbps or higher.

Plotting the histogram of the packet lengths during the video call shows statistically the anecdotal observation made by looking through the packet decodes. The RTS, CTS, QoS Data, Block ACK sequence with a 1400 byte UDP data burst is repeated throughout the video call. This packet sequence dominates the WiFi traffic during the video call. This is illustrated by the histogram shown in figure 16. Packet sizes are either less than 40 bytes or around 1400 bytes. The small byte size packets are signaling messages, RTS, CTS, and Block ACK. The packet sizes concentrated around 1400 bytes are video packets.

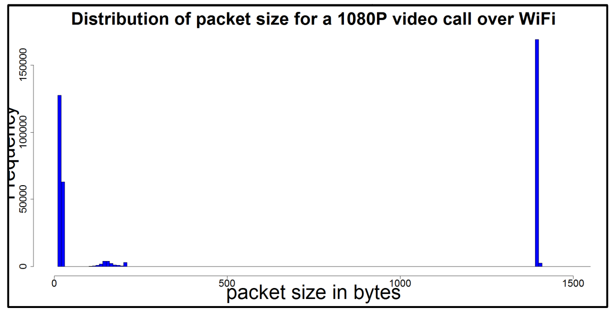

Fig. 16 Video call packet sizes are either small or large

Much of this analysis of the WiFi packets during a video call is focused on allowing the calculation of duty cycle, the percent of the time the application needs to use the RF spectrum. The reason that this is so critical is that WiFi uses unlicensed spectrum and thus any application must be judged based upon how well it will work while sharing the spectrum with other devices and applications. It is not a valid excuse for wireless LAN equipment and applications to claim that poor performance is due to a "noisy" environment. By "noisy" it is meant that other users of the spectrum are preventing the equipment or applications from working. However, equipment and applications using unlicensed spectrum must be designed to work in a shared spectrum environment. Users of unlicensed band equipment and applications must not expect performance levels that can only be realized with unshared spectrum. Even licensed band spectrum suffers considerable interference from adjacent cells and from spectral spillover from harmonically related or adjacent spectrum bands. So even licensed band equipment and applications must be designed to operate in the presence of fading and interference.

WIFI PACKET ANALYSIS OF A 720P CALL

A video call was set up with both callers having a 720P camera and display. One of the computers used a wireless home network connection. The wireless home network used

2.4 GHz channel 8 with a 20 MHz bandwidth and both the AP and the STA had 2 stream capability. The distance between the wireless access point and the wireless client station spanned about 36 feet, two floors and three walls of a residential home. The received signal strength level at the wireless client station ranged between -68 and -72 dBm. The video call lasted about 6 minutes and 30 seconds.

The video calling software reported call quality technical information. The video send stream was 1280 by 720 H.264 at 30 frames per second with a 1522 kbps data rate. The video receive stream was 1280 by 720 H.264 at 30 frames per second with a 1507 kbps data rate. The call technical information reported 0 relays indicating that the UDP traffic flowed directly between the two callers without intermediate nodes. The set up and data rate are shown in figure 17. The video call experience was excellent during this test. One video caller has an Ethernet 1 Gbps connection while the other video call uses a wireless home network connection with challenging RF signal conditions.

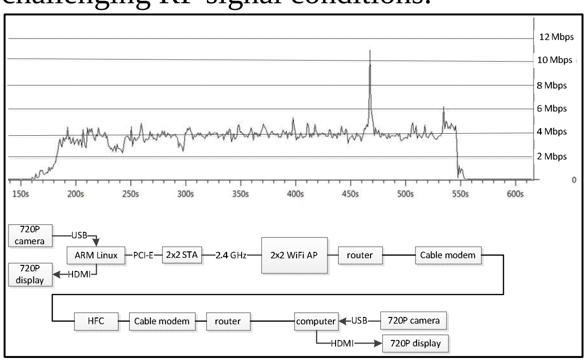

Fig.17 Video Call Set Up and Data Rate 720P 2.4 GHz.

During the video call an Airmagnet WiFi analyzer was used to capture the wireless local area network traffic. In all, 408,803 WiFi packets were captured and used for the statistical analysis of the call. The WiFi analyzer packet capture data file was saved as a Wireshark file and Wireshark analysis was used to create the IO data rate graph. The Wireshark data was then exported to a text file and a Perl program was written to extract and calculate a data array consisting of the burst time of the 408,803 packets. The R statistical programming language was then used analyze the distribution of the WiFi burst times.

The summation of the burst time of all the WiFi packets was 92.208012 seconds. Since the call lasted 6.5 minutes or 390 seconds, the percentage of time that the video call computer wireless station was either transmitting or receiving was 23.6%. In other words, the duty cycle of the 720P video wireless home network was found to be about 25%, one quarter of the time. If four such video calls were made utilizing the same wireless spectrum then we would expect conflicts due to 100% spectrum utilization.

The mean burst duration was calculated to be 225 µs. The median burst duration was found to be 40 µs indicating that many of the bursts were of short duration such as RTS, CTS, and Block ACK signals. The standard deviation of the burst times was 380 µs.

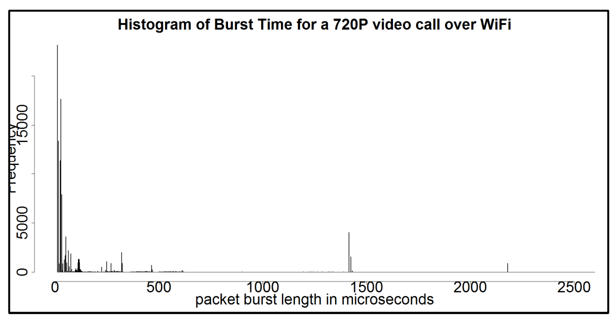

Fig 18. Histogram of the WiFi Burst Duration during a 720P video call.

The histogram of the burst durations is shown in figure 18. The bursts that last longer than 2 ms are beacons.

Figure 19 shows the histogram of packet burst time duration from 20 µs to 1 ms. The spreadsheet in Table 2 shows the percentage of packets for each possible data rate of transmission. With this histogram and spreadsheet it is easy to identify the main data rates used for sending video packets of about 1400 bytes. The three most prominent data rates are 13, 19.5, and 52 Mbps with burst durations of about 950, 640, and 260 µs, respectively.

Data Rate	Burst length	Burst Time	RX	TX	RX	TX
Mbps	bytes	microseconds	frames	frames	%	%
1	16	156	453	9,008	0.19%	5.31%
2	16	92	366	0	0.15%	0.00%
5.5	1495	2204	1	0	0.00%	0.00%
6.5	1495	1868	375	5,633	0.16%	3.32%
11	16	40	90,123	46,470	37.93%	27.40%
12	16	40	36,099	0	15.19%	0.00%
13	1495	948	1,159	27,950	0.49%	16.48%
19.5	1495	644	8,375	52,457	3.52%	30.93%
24	16	36	29,534	14,328	12.43%	8.45%
26	1495	488	7,615	10,199	3.20%	6.01%
39	1495	336	12,127	1,425	5.10%	0.84%
52	1495	260	50,095	1,041	21.08%	0.61%
58.5	1495	236	2	34	0.00%	0.02%
65	1495	212	17	49	0.01%	0.03%
78	1495	184	1,284	676	0.54%	0.40%
104	1495	144	1	338	0.00%	0.20%
117	1495	132	0	10	0.00%	0.01%
			237,626	169,618		

Table 2. 720P video call data rates and burst time

There are a couple points of interest in this analysis. First, although both the wireless access point and wireless station have two transmit and receive antenna chains and are

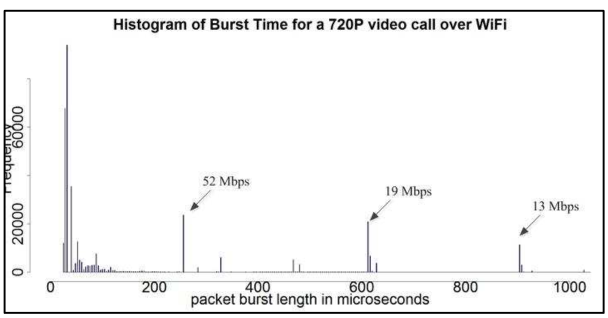

Fig. 19 Histogram of the WiFi Burst Duration up to 1 ms.

thus capable of dual stream operation, the data rate rarely goes above 65 Mbps and most video packets are being sent at a data rate lower than 65 Mbps. This is significant because a single antenna wireless station lacking dual stream capability will max out at 65 Mbps for 20 MHz channel width and normal guard interval. Under these circumstances, the single antenna client station is at no disadvantage compared with a multi-antenna client. In fact, with only one antenna chain the power consumption is reduced and there is less physical footprint to pick up on board interference. The late Steve Jobs was noted for his passion for simplicity and functionality. He demanded products that worked and were a pleasure to use. Long battery life and comfortable operating temperature trumped the fastest Mbps claim on the outside of the box. This is reflected in mobile products that for the most part use a single antenna design with 20 MHz channel width and normal guard interval.

The second thing to note is that this test set up is operating at the threshold of a successful 720P video call. A significant portion of packets are operating at 13 Mbps having burst duration of almost a millisecond. This is good enough for a 720P call and as we’ve seen only a quarter of the RF spectrum is utilized for this application, meaning that co-existence with other applications is reasonable. However, any lower modulation and coding schemes than this and the 720P video call will not work. Once operation goes below the 13 Mbps data rate bursts, the video calling software will reduce the video quality due to packet loss and jitter measurements. And this will be particularly noticeable if any competing traffic or applications are sharing the spectrum.

Video Call with a smart phone over a wireless home network

A video call was set up between a PC and a smart phone. The display of the smart phone had a 4.3 inch diagonal and the video camera was 1080P. The smart phone connected over WiFi 2.4 GHz to a home wireless gateway with integrated WiFi and cable modem. A speed test application run on the smart phone measured a latency of 29 ms, a download speed of 5294 kbps and an upload speed of 7968 kbps. The gateway and the smart phone were separated by 36 feet one floor and two walls.

The other end of the call was a PC with a 1080P video camera and display connected to the Ethernet interface of a router and a cable modem. Two different cable modems were used in this test so that the video call packets would have to traverse the HFC network to the CMTS. The same CMTS terminated both cable modems in this test. The data rate and block diagram of the test is shown in figure 20.

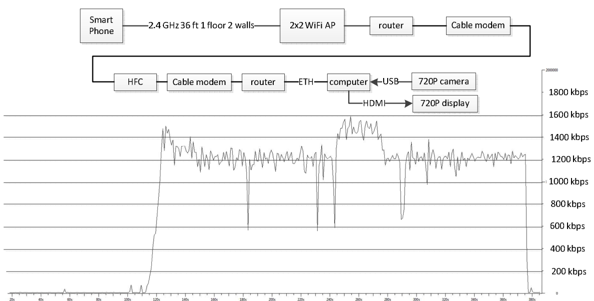

Fig. 20 Block Diagram and Data Rate of video call of smart phone with WiFi network connection.

The call technical information reported by the video calling software was monitored during the call. The number of relays was 0. The roundtrip time was 19 ms. The jitter was 69. The packet loss was 0.1%. The call lasted for 380 seconds or about 6 minutes.

The video send stream was 640x480 at 15 frames per second with H264 coding and 549 kbps bit rate. The video receive stream was 320x240 with H264 coding at 14 frames per second and a 605 kbps bit rate.

The number of packets captured for analysis was 60,348. The traffic protocol was UDP. The average data rate during the video call was 823 kbps and the average packet size was 648 bytes.

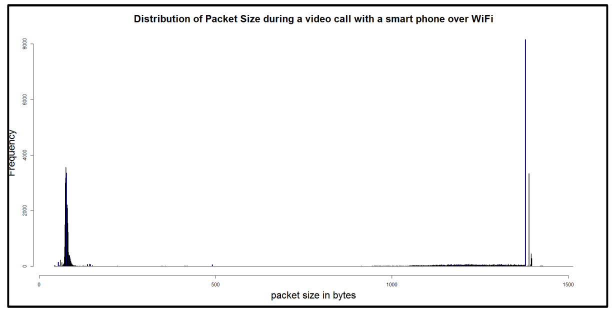

Fig. 21 Distribution of Packet Sizes during a video call using a smart phone with wifi network connectivity, x axis is packet byte size from 0 to 1500

Figure 21 shows the distribution of packet sizes during the video call using a smart phone with WiFi network connectivity. The packets are either very large or very small. The UDP video packets are typically about 1400 bytes whereas the WiFi signaling packets are typically less than 20 bytes in length. This explains the barbell type distribution of packet sizes.

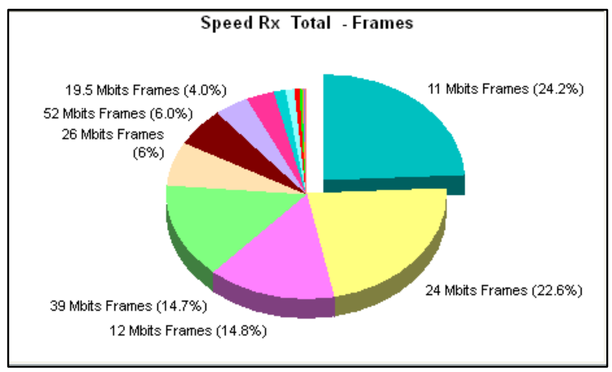

Fig. 22 Smart Phone over WiFi video call

Figure 22 shows the percentage of frame types with various data rates. A WiFi packet analyzer was used to create the pie chart. The majority of the packets had a data rate of 11 Mbps representing 24.2% of all WiFi packets sent. These packets are signaling packets, typically RTS,CTS, or Block ACK packets with short lengths of 16, 10, and 28 bytes respectively. 22.6% of the frames were 24 Mbps which are also signaling frames. The largest percentage of data carrying frames was the 39 Mbps frames representing 14.7 percent of the total number of frames. The 39 Mbps frames carry the large UDP video packets of about 1400 bytes of payload data. 14.8% of the frames were 12 Mbps. 6% of the frames were 26 Mbps. 6% of the frames were 52 Mbps. 4% of the frames were 19.5 Mbps.

During this video call using a smart phone with WiFi connectivity the WiFi analyzer captured WiFi network packets, the output was saved as a text file and a PERL program was written to calculate the burst duration of the 130,226 packets captured based upon the data rate and the byte size. The video call lasted 142.413 seconds and the period of time that the WiFi client was either transmitting or receiving was found to be 27,290,218 microseconds. Dividing the latter number by the former allows us to calculate that the utilization factor of the wireless spectrum during the video call was 19.2%. A histogram of the burst times is shown in figure 23. The predominate data rate of 39 Mbps for video packets of 1400 bytes which has a burst time of 336 microseconds is clearly indicated in the histogram. All in all it has been determined that a video call over a wireless home network using a smart phone has a data rate of 1 Mbps and uses up about one fifth of the wireless channel capacity.

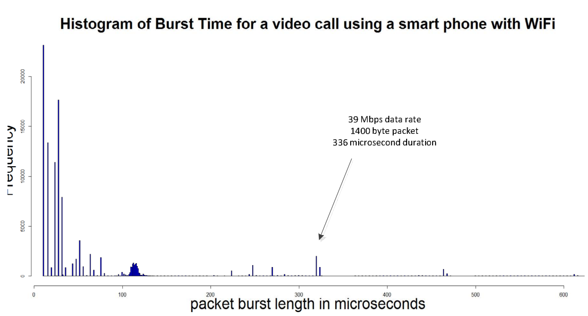

Fig. 23 Smart Phone over WiFi video call

Video Call over Cellular Wireless Networks

Video calls can be made over both 3G and 4G cellular networks. Here 3G networks refer to CDMA based networks and 4G networks refer to OFDM based networks since the characteristics pertinent to video calling varies considerably between these two multiplexing techniques. From a standards body standpoint, and from a service marketing standpoint, the use of the terms “3G” and “4G” is much more complex and nuanced and outside the scope of this paper.

Packet analysis was performed on a video call using a 3G cellular network lasting 1589 seconds or about 26 minutes. The video call quality was poor and the call dropped and re-established many times during the conversation. Still, the 3G end of the call was in the beautiful Florida Keys and the overall video calling experience was satisfying, refreshing to see a warm beach on a sunny day while huddled inside to avoid a cold grey Philadelphia winter. Video calling using a smart phone with a 3G data connection can be quite good at times as long as there is some tolerance for occasional disconnects, screen freezes, and fuzzy video.

Packets were captured with Wireshark on a PC with an Ethernet connection. The PC established a video call with another PC using a 3G cellular data card. The number of packets captured was 116,477. The average packet size was 295 bytes and the average data rate was 173 kbps.

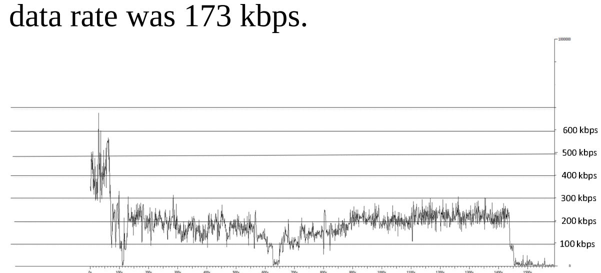

Fig. 24 Data Rate Measured During 3G video call

Figure 24 shows the data rate measured throughout the video call over the 3G cellular network. The data rate peaks at about 500 kbps, shows two steep drop offs where the call was lost, and otherwise runs at about 200 kbps.

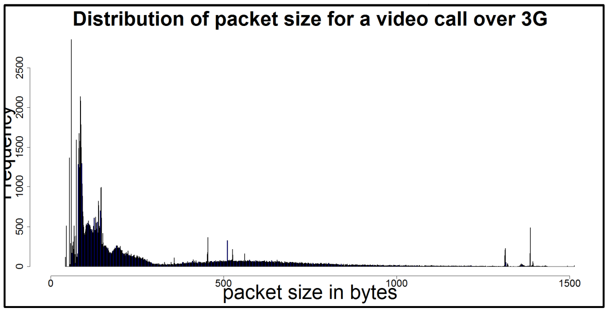

Fig. 25 Distribution of Packet Sizes of 3G video call.

The Wireshark packet analysis was exported to a text file, a PERL program was written to create an array of all the packet sizes for statistical analysis with the R statistical analysis tool. The resulting histogram is shown in figure 25 with the x-axis being the packet size in bytes ranging from 0 to 1500 bytes. By comparing the distribution of packet size between the 3G cellular network with that of the wireless home network, one notices that the packet sizes are generally much smaller when making a video call using the 3G network when compared to using a home WiFi network.

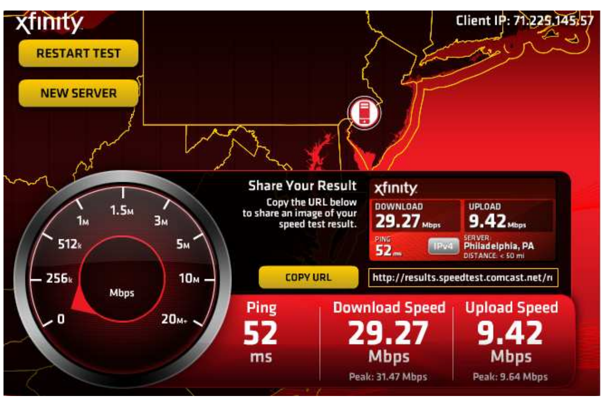

Fig.26 Speed test of 4G 700 MHz 10 MHz FDD pair

With the introduction of 4G networks having much higher data rates, and much lower latency and jitter, video calling over cellular networks will become better and more reliable. Like WiFi, 4G networks use OFDM which has a guard band in time to reduce inter-symbol interference as compared to a rake receiver or some type of adaptive equalizer used in CDMA networks. The adapter equalizer techniques used for single carrier wideband systems work well at times but require knowledge of the channel impulse response and so have difficulty under rapidly changing multi-path conditions. OFDM with a much simpler guard time inter-symbol interference mechanism can work even under rapidly changing multi-path conditions.

As the channel width increases the impulse response gets more complicated, requiring more taps for an adaptive equalizer and more calculations to respond to changes in multi-path conditions. This limits the channel width of CDMA based systems. The channel width used in the 3G video call of figure 24 and 25 has a 2 MHz channel width using CDMA. There are also 3G CDMA networks with 5 MHz channel width.

4G OFDM systems can operate at increased channel widths of 10 MHz, which gives them higher data throughput. Figure 26 shows the speed test results for a 4G network operating in the 700 MHz spectrum band with two frequency division duplexed, FDD, 10 MHz channel width signals. The download data rate is 29 Mbps and the upload data rate is 9 Mbps with 52 ms latency. These data rates, if maintained throughout the course of the call, are sufficient for 1080P video calling. One caution, cellular networks tend to be used in cars, buses, trains, or even when walking around and while moving throughout a geographical area the data rate will vary significantly and even switch from 4G to 3G and 2G coverage areas. So it is unlikely to always maintain these speeds while moving. In the area where video call testing was performed for this paper, 4G coverage was not available so video call testing and analysis was performed using a 3G network.

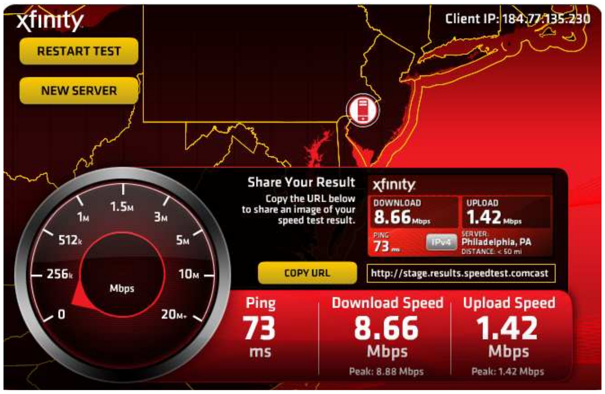

Fig. 27 4G network 2.5 GHz, 10 MHz TDD

Figure 27 shows the speed test results of a 4G network operating in the 2.5-2.7 GHz spectrum band with a 10 MHz channel width using time division duplexing, TDD. The measured upload speed was 1.4 Mbps and the download speed was 8.6 Mbps with 73 ms latency. While the upload data rate was not high enough to support a 1080P video call, it was close to the 1.5 Mbps upload speed required for a 720P video call. The upload data rate of 1.4 Mbps is enough for a 500 kbps video send stream of a smart phone video call and the 8.6 Mbps download data rate has lots of room to support a 500 kbps receive video stream from a smart phone video call.

Figure 28 shows the parameters for the speed test results shown in Figure 27. The center frequency of operation is 2.647 GHz. The received signal strength is a very high -46 dBm indicating very good RF signal conditions and probable operation in close proximity to a base station. The carrier to interference and noise ratio was 21 dB. The transmit power was -19 dBm, the transmit power can go up as high as +20 dBm if the attenuation of the RF signal to the base station is high.

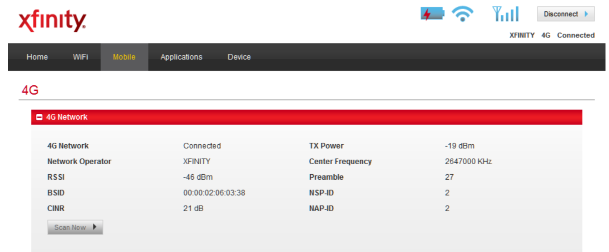

Fig. 28 4G Network 2.5 GHz, 10 MHz TDD

While 4G networks have the technical capability to make video calls under good RF conditions, will the costs impede usage? Figure 29 shows some calculations that translate a typical 4G data plan into some metrics familiar to many who in the past have bargain shopped for long distance plans based upon cents per minute or cellular phone plans based upon monthly minutes of talk time. The plan analyzed is a cellular data plan with 5 GB for $50 per month. If a video call is assume to have a data rate of 3 Mbps, the data rate measured for a 720P video call, then a video call is 21 cents per minute. At one point in time 10 cents a minute for a long distance plan seemed like a good deal. Cellular data plans with monthly minutes of talk time tend to be priced about 9 cents per minute. In terms of monthly talk time if one was to switch from voice calling to video calling, $50 per month with a 5 GB data cap gives one 238 minutes

$$bit := m \qquad Mbit := 10^{6}\ bit \qquad byte := 8\ bit \qquad GB := 2^{30}\ byte$$

$$MB := 2^{20}\ byte \qquad month := \frac{yr}{12}$$

$$GB = \left(1.074 \cdot 10^{9}\right)\ byte$$

$$monthly\_fee := 50\ \frac{\text{¤}}{month} \qquad monthly\_data := 5\ \frac{GB}{month}$$

$$video\_calling\_data\_rate := 3\ \frac{Mbit}{s}$$

$$video\_call\_minutes\_per\_month := \frac{monthly\_data}{video\_calling\_data\_rate}$$

$$video\_call\_minutes\_per\_month = 238.609\ \frac{min}{month}$$

$$video\_call\_cost := \frac{monthly\_fee}{video\_call\_minutes\_per\_month}$$

$$video\_call\_cost = 0.21\ \frac{\text{¤}}{min}$$

Fig.29 Calculating the cost of a 4G video call of talk time with video calls at 3 Mbps. Voice cellular plans in this price range tend to offer about 450 minutes of talk time. So with these parameters, making video calls rather than voice calls tends to cost about twice as much. Of course, the assumed bit rate of the video call is the critical parameter. If the video calls were all 10 Mbps 1080P then it would be very expensive, 70 cents per minute and only 70 minutes of monthly talk time. However, on the other hand many folks may be quite content with making video calls on the road with a smart phone operating at 1 Mbps, in this case the cost per minute is 7 cents with 715 minutes of monthly talk time. These last numbers are roughly equal to the cost of cellular voice calls. So if you have a smart phone and a 4G data plan, live it up, make a video call instead of a voice call.

CONCLUSION

Many things have come together recently to encourage the use of video calling. Broadband connections in the home are faster than ever. Many homes have wireless home networks to connect mobile devices. In the past providing Internet connectivity to your television may have been inconvenient due to the lack of a nearby CAT-5 outlet. The wireless home network takes away the inconvenience. More and more people carry smart phones with WiFi and 3G or 4G network connectivity and these smart phones have front and back cameras and video calling application software.

Tests of video calls have shown that a 1080P video call can run at a symmetrical 10 Mbps, with the video send stream at 5 Mbps and the video receive stream at 5 Mbps. An excellent quality 720P video call on a large screen television set can run at 3 Mbps total data rate. 3-way video calls tend to run at about 2 Mbps. Smart phone video calling with a 4.3 inch diagonal display runs at about a 1 Mbps data rate over a home WiFi network. Video calling using a 3G cellular network runs at about 200 kbps with lower video quality and reliability.

Video signals are sent in packets of about 1400 bytes. Wireless home networks supporting video calls tend to have very concentrated packet size distribution around 1400 bytes and 20 bytes, representing the video packets and signaling packets, respectively. A typical packet sequence of a video call over WiFi lasts about 250 μs and consists of a request to send signal, a clear to send signal, the data packet of 1400 bytes, and a block acknowledgement signal.

Tests were performed of a 720P video call and a 1080P video call whereby all of the WiFi packets were captured. The captured packets contained information on byte size on the wire and data rate of the modulated burst. Accounting for the preamble length, the time length of each packet transmission was calculated and statistically analyzed. With the duration of the video call and the transmission time of each packet during the call determined, the percentage of time that the wireless home network was used by the video calling application was determined. For a 1080P video call under ideal RF conditions, the duty cycle was found to be 25%. For a 720P video call under threshold RF conditions the duty cycle was 20%. This indicates that most wireless home networks could support no more than 4 to 5 simultaneous video calls and that even during a video call over a wireless home network there is still over 75% of the capacity available for spectrum sharing.

Finally, the distribution of packet sizes of a video call using a 3G network was measured and analyzed. The packet size distribution shows that packet size in general was not as large when making a video call over a 3G network. The video calling application adjusted to the higher latency and packet loss of the 3G network in order to make the call while sacrificing quality. The speed of 4G networks was measured and reported indicating that 4G networks do support the data rates required for high quality video calling, at least under ideal RF conditions. The cost of video calls on 4G networks was analyzed and it was found that with today's pricing plans, video calls of very high quality are more expensive than voice calls but not prohibitively so, while lower quality video calls on a smart phone screen today cost about the same as most common voice plans.

Wireless network connectivity is a crucial factor in encouraging the use of video calling. The signal strength is a critical indicator, wireless home networks should have signal strength indication of -60 dBm or higher for reliable video calling. Signal strength of -70 dBm for a wireless home network connection was found to be at the threshold of operation for a successful video call. The use of 5 GHz band can work better than 2.4 GHz band but only at close proximity. It was found that during a video call with a 3x3 AP and 3x3 client at 5 GHz in close proximity that 3 stream operation was very rare. It was also found that during a video call with a 2x2 AP and 2x2 client at 2.4 GHz at a 36 foot AP to STA separation distance that 2 stream operation was very rare.

REFERENCES

[1] Dorros, Irwin, "Picturephone", Bell Laboratories Record, Vol. 47, Number 5, May/June 1969, pp. 136-141.

[2] Urban, D., Albano, C., Devotta, D., "Delivering DOCSIS 3.0 Cable Modem Speeds over the Home Network", SCTE Cable-TEC EXPO'10, October 2010.

[3] Urban, D., Albano, C., Ong, I., Gilson, R.,"Designing a reliable wireless home network in a residential environment to optimize coverage and enhance application experience", SCTE CABLE-TEC EXPO '11, November 15-17 2011 Atlanta Georgia.

[4] IEEE Std. 802.11n-2009, "Amendment 5: Enhancements for Higher Throughput.

ABBREVIATIONS

AP Wireless local area network access point
STA Wireless local area network client station
OFDM orthogonal frequency division multiplexing
GI guard interval for OFDM
CDMA code division multiple access
HFC Hybrid Fiber Coaxial Cable network architecture
CM cable modem
RTS request to send WLAN signal
CTS clear to send WLAN signal
Block ACK Block Acknowledgement WLAN signal
WLAN Wireless local area network
LTE Long Term Evolution 4G cellular network
WiMAX type of 4G cellular network
4G OFDM based cellular network
3G High speed CDMA cellular network
RTT round trip time in ms
HT High Throughput WiFi mode
MIMO multiple input multiple output antennas
IFFT inverse fast Fourier transform
FFT fast Fourier transform
TDD Time Domain Duplexing
FDD Frequency Domain Duplexing

ACKNOWLEDGEMENTS

The author wishes to thank colleagues at Comcast and especially Chris Albano, Doug Jones, Erami Botchway, Sam Chernak, Chris Bastian, Jorge Salinger, and Neil Spence for help and support in writing this paper.