EMPOWERING HD AND 3D VIDEO STREAMING

Benny Bing (benny@gatech.edu)
Georgia Institute of Technology

Abstract

This article presents methods to improve the efficiency and performance of streaming high-definition 2D and 3D compressed videos. Two key methods involve traffic shaping and buffer dimensioning. It will be shown that these methods reduce bit rate variability, which will in turn, minimize packet losses due to buffer overflows at the receiving device. We will also show that with efficient multiplexing and aggregation, further performance gains may be achieved.

1. INTRODUCTION

Streaming live and on-demand digital video content over the Internet, and in telecommunications and broadcast networks, is becoming prevalent. Streaming variable bit rate (VBR) video traffic is a special challenge due to the high dynamic range of the frame sizes that results in high bit rate variability. This is especially so for HD and 3D videos. As an example, Figure 1 shows the high and variable encoded rates for a 720p 3D video compressed using VC-1. Shaping the traffic to reduce the peak rates will minimize packet losses due to buffer overflow at the receiver.

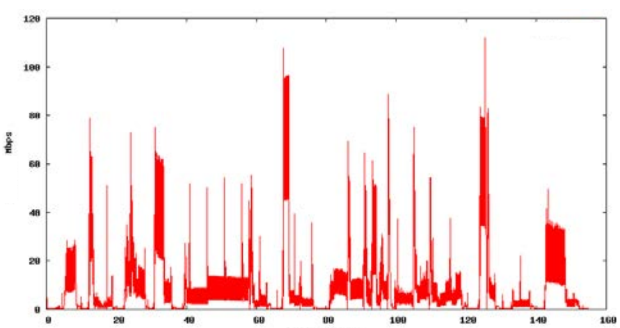

Figure 1: Variable bit rates for 3D 720p VC-1 video.

2. IMPACT OF PLAYER'S BUFFER SIZE

Figure 2 shows the impact of the player's buffer size on the TCP streaming throughput for a 720p H.264 video. A larger buffer of 8 Mbytes allows more information to be stored, hence the transmission can be completed at an earlier time than the case with a 1 Mbyte buffer. Note that the advertised receive window size in TCP may throttle the throughput to a smaller value when the buffer has received sufficient information. This is because the sender will take the minimum of the congestion window (which attempts to avoid congestion) and the receive window, when deciding on the appropriate window size to use. When some of the frames are played out, more buffer space is released, and the receive window will increase again. Figure 3 shows the impact of shaping the streaming throughput to a rate limit of 1.5 Mbps. As expected, the video file is received later than the unlimited case and the player with a larger buffer size achieves more efficient bandwidth utilization.

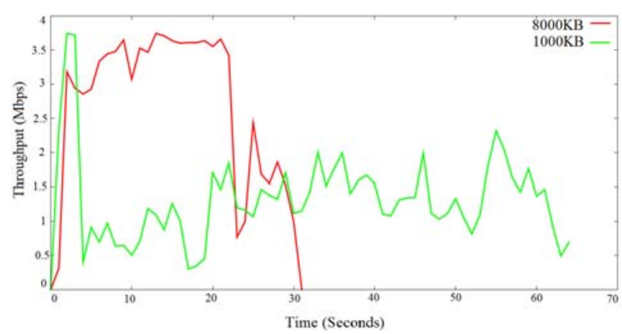

Figure 2: Impact of player buffer size on TCP streaming throughput.

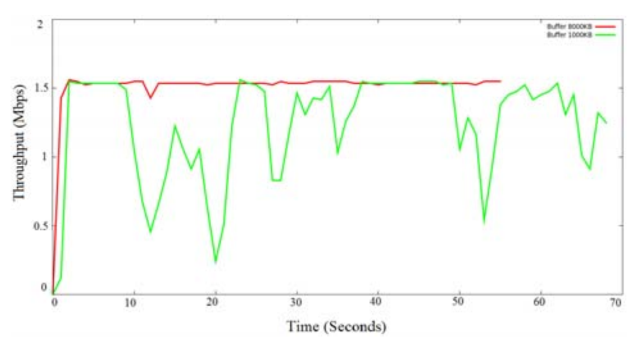

Figure 3: Impact of buffer space and shaping on TCP streaming throughput.

3. IMPACT OF TRANSPORT PROTOCOLS

The performance of traditional TCP and TCP with SACK is shown in Figures 4 and 5. The player's buffer size was set to 600 Kbyte. The theoretical average rate is computed by taking the video file size and dividing by the duration of the video. The initial rate is usually high because the player attempts to buffer as much data as possible for playback (including the first frame, a large I-frame), and thus advertises a large receive window. Subsequently, this advertised window gets throttled to a reasonable value. For the Avatar 720p 3D movie, the initial high rate can be attributed to the nature of the video content (high action at the start) and the player. From Table 1, it is clear that TCP with SACK reduces the overheads and hence, the average streaming rate. This is because TCP with SACK employs aggregated or block ACKs, which reduces the overheads of sending individual ACKs, thereby enabling multiple packet losses to be handled more efficiently.

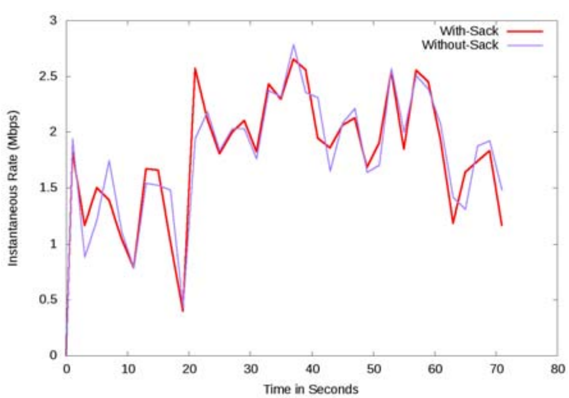

Figure 4: Streaming rates for a 1080p H.264 video.

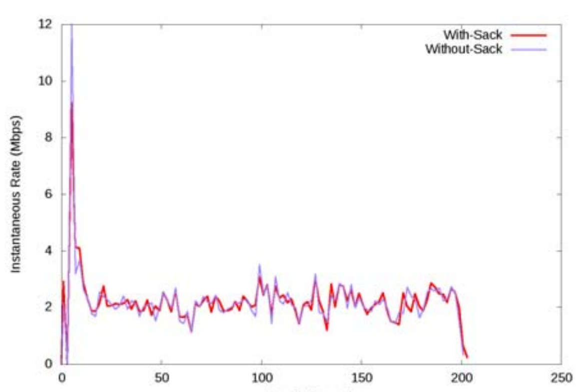

Figure 5: Streaming rates for a 720p H.264 3D video.

Table 1: Comparing TCP overhead (average rates).

	1080p 2D	**720p 3D**
TCP with SACK	2.011	2.318
TCP without SACK	2.014	2.326
Theoretical	1.836	2.111

Figures 6 and 7 show the streaming of the 720p 3D trailer using UDP. In general, UDP incurs less overheads than TCP since it is a unidirectional protocol. Unlike TCP, there is no congestion flow control in UDP, hence the raw UDP sending rates can be very high (a few orders of magnitude higher than TCP) and the video gets transferred in a very short duration. This may result in unacceptably high loss rates due to network congestion and receiver buffer overflow. These losses cannot be recovered in native UDP. A solution to manage this problem is to send the video at the frame rate that the video is meant to be played back (in this case, 30 Hz). The resulting sending rates are shown in Figure 6. Alternatively, progressive streaming can be employed to shape the rate variation (Figure 7). In both cases, the entire video gets transferred according to its duration (208s). For progressive streaming, the shaping threshold is set to 2.13 Mbps, thus proving that the UDP overheads are lower compared to TCP (Table 1).

3. PEAK TO AVERAGE RATE

The peak to average rate (PAR) normalizes the actual variation of the VBR video rates. This is computed by selecting the instantaneous rate of the compressed video and dividing by the average rate within a predefined interval. For example, if the video frame rate is 25 Hz, then the frame interval is 40 ms. The size of the frame divided by the frame interval gives the instantaneous rate. The predefined interval is chosen to be the duration of the entire video, and the instantaneous rate for each frame interval is averaged over this period, giving the long-term average PAR. Figures 8 and 9 illustrate the instantaneous rates and the long-term average PAR for a 720p H.264 movie trailer.

Note that the peak rate for the 720p VBR H.264 video is over 25 Mbps. Shaping the traffic to a lower rate results in additional buffering delay that has to be accommodated. As shown in Figure 10, choosing a low shaping threshold can result in significant delay. As we will see in the next section, a better way is to aggregate multiple video streams to make more efficient use of the channel bandwidth.

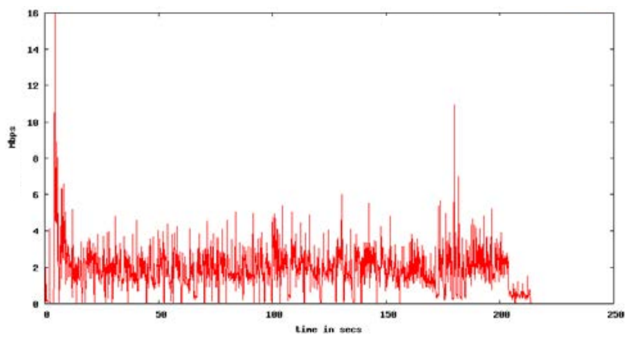

Figure 6: Instantaneous UDP rates for streaming the 720p H.264 3D video at 30 Hz.

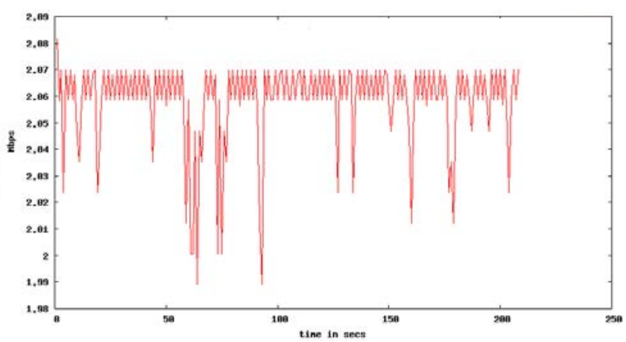

Figure 7: Shaped UDP rates for streaming the 720p H.264 3D video.

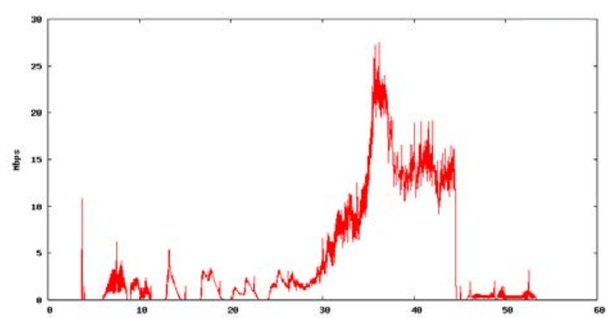

Figure 8: Instantaneous rates for a 720p H.264 video.

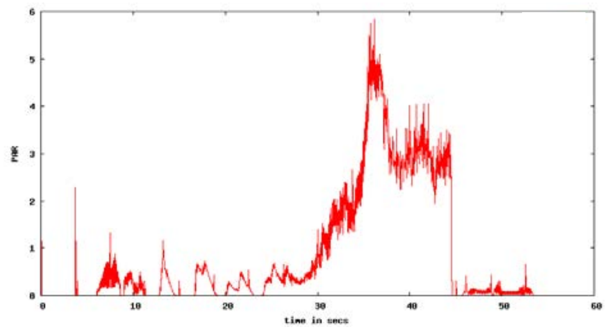

Figure 9: PAR for a 720p H.264 video.

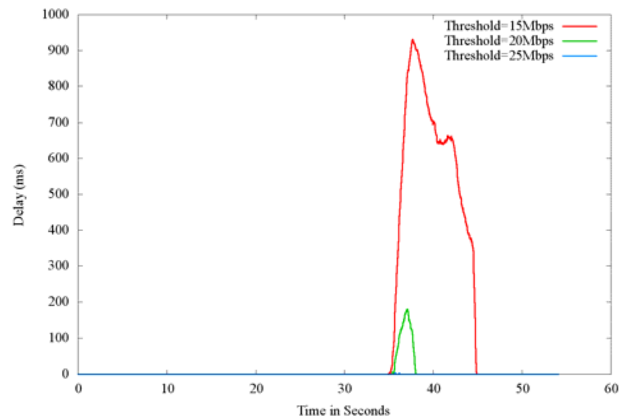

Figure 10: Shaping thresholds versus buffering delay.

The PAR variation is very high. A PAR of 1 is desirable because it achieves perfect utilization of the link bandwidth. A PAR value below 1 implies under-utilization whereas a PAR greater than 1 requires more buffering or bandwidth to accommodate the peak rates of the VBR video. By appropriately shaping the video traffic, a PAR close to 1 can be achieved. Figure 11 shows the streaming of another 720p H.264 video with a fixed shaping threshold. A PAR close to 1 is achieved. Since the video was streamed in real-time using TCP and the duration of the video was not known beforehand, a running average method was used to compute the PAR. The running average is computed by averaging the instantaneous rates since the beginning of the video play back.

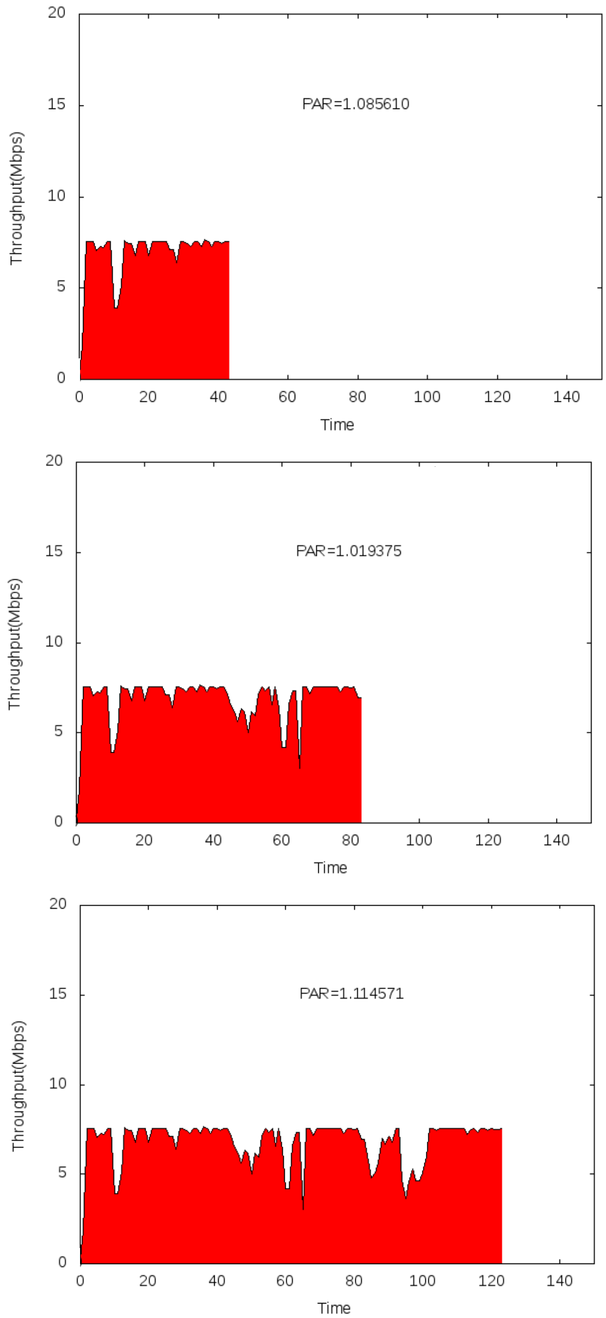

Figure 11: Evolution of PAR for live TCP streaming of a shaped 720p H.264 video.

4. MULTIPLEXING OF VIDEO STREAMS

It is common for a video headend or server to deal with multiple streams. When statistical multiplexing is applied to multiple VBR compressed videos at the headend or server, it can exploit the inherent variations in the instantaneous bit rates and increase the number of video streams within a fixed channel bandwidth while keeping the picture quality constant. For example, if one stream is demanding high bit rate, it is likely that other streams have capacity to spare. A large number of aggregated streams tends to "smooth" to a normal distribution (based on the central limit theorem). Unlike per stream buffer-based traffic shaping, statistical multiplexing introduces minimal delay.

The data rates for a MPEG stream can vary quite dramatically depending on the video content. As shown in Figure 12, the video resolution also plays an important part in the frame size distribution (and hence the data rates). The 480p Dell video contains very fast scene changes and the peak of the frame sizes occurs in the 170 Kbyte range. The 1440 × 1080 Terminator-2 trailer is fast action and the peak of the frame sizes occurs in the 340 Kbyte range. Compare with the slower motion 1080p FCL movie (only 5 scene changes) where the peak of the frame sizes occurs in the 480 Kbyte range. In addition, this video exhibits the broadest range of frame sizes. However, there is no strong correlation between the frame size distribution and the video duration: Dell (75.3 s), Terminator-2 (125 s), FCL (72.2 s).

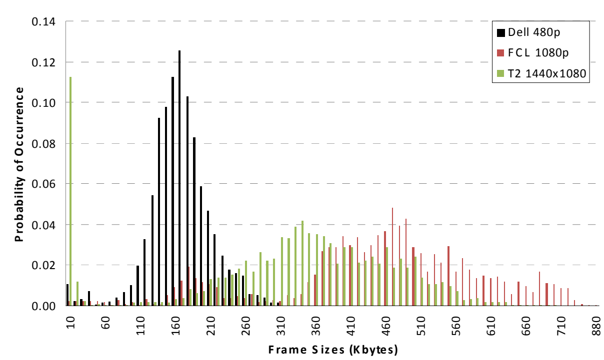

Figure 12: Frame size variability of H.264 videos.

Many compressed videos exhibit the long-range dependent (or long-tail) traffic characteristic. Because of this dependency, the video traffic tends towards clustering and becomes less predictable as the number of streams increases (which is in contrast to Poisson distributions that become smoother as the aggregation volume increases). To illustrate this phenomenon, we multiplex 19 videos and 38 720p H.264 movie trailers with different content. As shown in Table 2 and in Figures 13 and 14, the standard deviation of the multiplexed rates is almost doubled when the number of multiplexed videos is increased two-fold, thereby proving the increased variability for a higher number of multiplexed streams. Thus, for multiplexed video streams, the buffers need to be larger to accommodate more extreme traffic-burst scenarios and traffic shaping may be needed. However, the standard deviation of the long-term average PAR reduces. This is because the aggregated average rate for 38 streams is larger than 19 streams, and this in turn, masks the effect of the overall variation to some extent. The average rate for the 19 and 38 streams remains about the same – 4.7 Mbps. Note that for 38 streams, a reasonable shaping threshold of say 250 Mbps results in minimal delay and yet, accommodates an average rate of 6.6 Mbps per stream, far lower than the 28 Mbps peak rate in Figure 8. Similarly, for 19 streams, a shaping threshold of 140 Mbps results in a delay comparable to 38 streams, giving an average rate of 7.4 Mbps.

Table 2: Standard deviation for instantaneous rates and PAR of multiplexed videos.

Number of Streams	Instantaneous Rates	Long-Term Average PAR
19 streams	28.797	0.3006
38 streams	50.391	0.2501

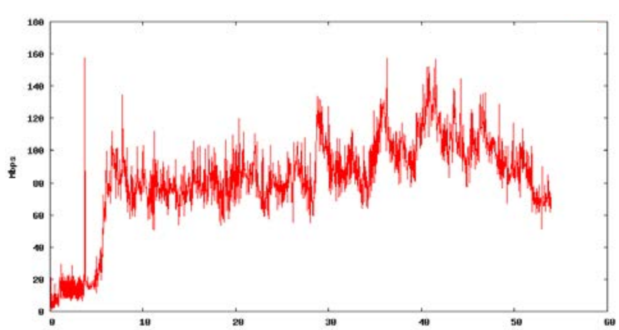

Figure 13: 19 multiplexed 720p H.264 videos.

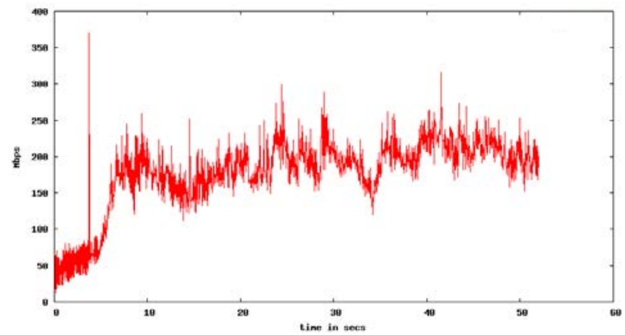

Figure 14: 38 multiplexed 720p H.264 videos.

To sum up, the bandwidth efficiencies that can be achieved using CBR and VBR 720p video multiplexing are shown in Figure 15. With channel bonding, higher efficiencies tend to be possible but not guaranteed. This is because with more streams, the standard deviation of the overall instantaneous rate becomes higher. To attain the same shaping delay as the lower number of streams, some bandwidth efficiency must be sacrificed. Alternatively, more aggressive shaping can be employed but this results in higher delays.

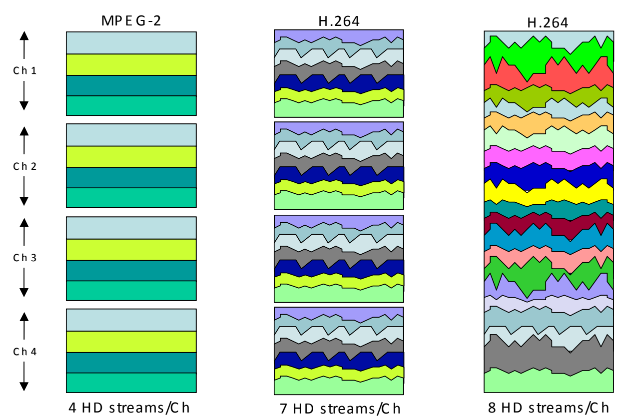

Figure 15: Efficiencies of 720p video multiplexing with and without channel bonding

5. VIDEO CONTAINER FORMAT

We now evaluate efficiencies of the MP4 video container format, which is widely used by online video portals. As can be seen from Table 3, the overhead for encapsulating a H.264 video in MP4 is insignificant, well below 0.01%, roughly 12 bytes per video frame. The MP4 container also incurs less overheads than the MPEG-2 transport stream

(TS) container, which has been wide used in many cable systems (Table 4). The difference in overheads grows as the video file size increases but the average percentage increase is in the region of 4%.

Table 3.6: Overheads for mp4 encapsulation (no audio).

Video	H.264 File Size (Mbyte)	MP4 File Size (Mbyte)	Overheads (Kbyte)
75s Dell, 480p	373.157	373.185	28
72s FCL, 720p	387.224	387.246	22
72s FCL, 1080p	772.062	772.085	23
596s BBB, 1080p	4,213.150	4,213.308	158

Video in 1280 × 544p	TS File Size (Mbyte)	MP4 File Size (Mbyte)	Difference (Mbyte)
54s The Dark Knight	33.395	32.094	1.301
64s The Hangover	45.736	44.055	1.681
73s Fred Claus	51.796	49.936	1.860
83s Night at the Museum	59.716	57.514	2.202
86s Speed Racer	62.907	60.595	2.312
106s 300 Video	54.579	52.479	2.100
128s Star Wars Clone Wars	94.247	90.910	3.337
137s The Astronaut Farmer	94.140	90.663	3.477
143s 10000 BC	98.045	94.540	3.505
141s Brothers Bloom	100.209	96.505	3.704
150s Transformers	110.419	106.415	4.004
191s Tetro	141.744	136.442	5.302

6. CONCLUSIONS

In this article, we have illustrated how HD and 3D VBR TCP/UDP video streaming can be enhanced via appropriate buffer size dimensioning, traffic shaping, and the use of statistical multiplexing to conserve bandwidth for aggregated video streams. We advocate the use of the MP4 container format for improved bandwidth efficiency. We have shown that the use of channel bonding may not improve the multiplexed bandwidth efficiency significantly, when compared to the single channel case. This is due to the long-range dependent characteristic of compressed VBR videos, which leads to an increased variability of the unshaped instantaneous rates when a higher number of streams are multiplexed. Our ongoing work focuses on deriving appropriate formulas for estimating the shaping thresholds and performing a more in-depth analysis of the long-term dependency of the HD and 3D VBR videos. Our longer-term research work focuses on enhancing the efficiency of the DOCSIS PHY layer via orthogonal frequency division multiplexing.

ACKNOWLEDGMENTS

The author is grateful to the support of Cox Communications.

REFERENCE

[1] B. Bing, *Broadband Video Networking – Empowering the HD and 3D Generation*, Artech House, Sept 2010.

BIOGRAPHY

Benny Bing is a research faculty member with the Georgia Institute of Technology since 2001. He has published over 80 technical papers and 11 books, and holds 5 pending patents in the areas of video, gesture, and bandwidth management technologies. His publications have appeared in the IEEE Spectrum and he has received 2 best paper awards. In early 2000, his book on wireless LANs was adopted by Cisco Systems to launch Cisco's first wireless product, the Aironet Wi-Fi product. He was subsequently invited by Qualcomm and the Office of Information Technology to conduct customized Wi-Fi courses. Other books were reviewed extensively by IEEE Communications Magazine (twice), IEEE Network as well as the ACM Networker. He is an editor for the IEEE Wireless Communications Magazine since 2003, where he also heads a special section on Industry Perspectives. He has guest edited for the IEEE Communications Magazine (2 issues) and the IEEE Journal on Selected Areas on Communications. All 5 of his online IEEE wireless tutorials have been sponsored by industry with one tutorial sponsored twice. In October 2003, he was invited by the National Science Foundation to participate in a workshop on Residential Broadband. He also led a team that received the National Association of Broadcasters (NAB) Technology Innovation Award in 2010. The award recognizes organizations that bring technology research exhibits and demonstrations of exceptional merit to the NAB Show. He is the founder of a video startup focused on enhancing and delivering next-generation video entertainment. He is a Senior Member of IEEE and an IEEE Communications Society Distinguished Lecturer.