Introduction To MPEG Video

For The Digital Network Professional

By Rich Mavrogeanes

Okay, so you know how to install and run a network. You can install a NIC in your sleep, routers don't scare you, you've migrated to a fully switched network, and when you think about ATM you don't assume it's a cash machine.

You're a data networking professional. You've been satisfied just to make the network work – despite all the changes and new requirements. You hear users clamor for more bandwidth all the time, and you have to do more with less. And now they want to add WHAT to your network? Are they kidding?   Won't video just kill the network? Better hold on, because video is coming to your network, and you'll want to speak the language.

You don't need to know how a network operates to use it, and you don't need to know how MPEG works to deploy TV-Quality video in a network. But it is interesting to know how it works and how to avoid network issues.

Just Another Class of Data?

From a network perspective, digital video is indeed "just data", but data that needs fairly reliable transport. Using IP Multicasting, the overall impact of adding video to a network can be minimal. Using ATM, the impact is close to zero. But this paper does not address those networking issues – you already know digital networking. This paper will help you to understand basic video concepts from a data networking perspective.

Video

Video is nothing more than a series of still images (frames) displayed fast enough that the human eye cannot detect flicker. There are three important characteristics of video:

- Resolution – how much information is in each picture
- Frame Rate – how often is a new picture displayed
- Color – how much color information is present

Video Resolution

Resolution is usually measured in horizontal and vertical pixels, where a pixel is a "Picture Element". As a data professional, you can easily become confused because in the computer world, resolutions are typically 640h x 480v to 1024h x 768v. In the TV world, it is 720h x 480v and 352h x 240v.   Obviously, the higher the resolution the more information is required. And if we are talking data communications, higher resolution means that more information has to arrive at a destination in one second than lower resolutions. Hence higher resolution means higher transmission bandwidth.

Copyright VBrick Systems, Inc.

Frame Rate

The frame rate is simply the number of pictures (frames) displayed in one second. Motion picture films, for example, are projected at 24 frames per second. North American television (NTSC) is displayed at 29.97 frames per second (we'll call this 30 fps), and European (PAL) video is at 25 fps. A networking professional can get confused here, since most computer terminals "refresh" the screen much faster – like 72 Hz or more. If displaying video on a computer, the refresh rate and the frame rate will certainly be different.

Color

This is the most confusing part of digital video, because there are so many possibilities! Video is made up of what is called "Luminance", which is the black-and-white or contrast portion of the picture, and the color component, which is, called "Chrominance". The network professional can think of color similar to the way we think of video resolution – more color requires more bandwidth. Consider a common digital color scheme where we use each of the primary colors[1]. If each color can be represented by 256 levels, than we can generate 16.8 million different colors (256 Red x 256 Blue x 256 Green). To support this level of a color, you need eight bits per color, and there are three colors. So it's 8 Red + 8 Blue + 8 Green – and we call this "24-bit video".

In the non-computer video world, color is represented by what is called "YCrCb". "Y" is luminance, Cr is "Chrominance-Red" and Cb is "Chrominance-Blue". This method for representing color takes advantage of the fact that the human eye does not detect color in the same way as it detects contrast. The human eye has many more "rods" which detect black-and-white images than it has "cones" that detect color. For the networking professional, what's important is that color information is present and is a key part of the video.

Analog Video

Video is typically analog at its source and destination. It might be a camera, the output of a TV tuner, a VCR, or a DVD player. Video is typically delivered on shielded coax cable terminated by RCA jacks or BNC jacks. The longer the analog cable, the more the loss. Analog video should not be confused with cable TV, which is RF modulated with analog video.

Make it Digital, Please!

So if we take analog video and digitize it, we end up with about 160[2] Mbps. How do we get there? Well, if the video is standard NTSC, it is 720h x 480v pixels. Multiple that times 2 bytes per pixel, then multiple that times 30 frames per second. Without compression, a single video stream would require more than the capacity of an OC3, and would swamp any corporate network.

[1] Primary colors for light are Red Green, Blue; Primary colors for pigments (paint) are Red, Yellow, Blue.

[2] CCIT-601 uncompressed video is 270 Mbps, using 10-bit sampling (we assuming 8 bit here).

Copyright VBrick Systems, Inc.

Resolution Compression
To reduce the bandwidth necessary for transmission, the video must be compressed. This is achieved using industry-standard MPEG. The first step in the process for MPEG-1 is to reduce the resolution from 720h x 480v to 352h x 240v. The reduced resolution results in a 4:1 reduction in bandwidth, since we halve both the horizontal and vertical resolution. The original 720h x 480v video size is far beyond what most monitors can display, and only the most expensive studio cameras can support. This has nothing to do with the actual size of the video when it is displayed, as it will still fill the screen. MPEG-2 does not reduce the resolution. The original resolution is called ITU-T 601 (CCITT 601), and is used by video professional who work with high resolution video so that multiple generations of editing and transmission do not degrade the quality beyond acceptable limits.

The Block Party
When you hold a motion picture film up to see each frame, you realize that only small regions of the picture change from frame-to-frame. MPEG breaks areas of each frame into 8 x 8 pixel blocks, and performs a mathematical function called Discrete Cosine Transform (DCT) to each block. DCT packs more of the information that a human detects into a tighter spot. For example, if the data pattern for a block looked like a string of 1's and 0's, after DCT, all the 1's would be bunched together and the more random 1's would not represent very important information. This is good, because long strings of 1's are easy to compress and if we throw away some of the more random 1's the human eye won't notice.

Round If Off
The next step in the MPEG compression process called "quantization". After DCT, the data in each block is rounded. For example, "1.1" becomes a "1" and "5.6" becomes a "6". Rounding reduces the amount of data and is the reason MPEG is considered a "lossy" compression method because once the rounding has occurred, the original accuracy can never be restored. For the data professional, the idea of "lossy" compression may be new, since video compression is not like putting data in a zipped file.

Remove Redundant Data
The data is subjected to a process to remove redundancy. For example, it is not necessary to transmit a string of twenty 1's – rather it is faster to send a single symbol that represents twenty 1's. This is accomplished for each block.

Motion Vectors
Now that we've got a series of blocks, each fully compressed, MPEG can do some interesting things. MPEG breaks the picture into a new set of blocks called "Macroblocks" which are 16 x 16 pixels. If there is a difference between the last frame and the current frame, MPEG only needs to move those blocks to a new location on the current frame (rather than sending a whole new picture). This important MPEG feature results in an

Copyright VBrick Systems, Inc.

additional bandwidth compression of up to 4:1, reducing the needed bandwidth substantially.

The World According to GOP

MPEG uses a series of I, B, and P frames which make up a "Group of Pictures".

An "I" Frame is made up of one video frame and it stands all by itself. A "P" frame is a predicted frame, and is based on past I frames. A "B" frame is a bi-directional frame made up of information from both "I" frames and "P" frames. The basic idea is that since it is highly likely that the "next" frame of any video will be very much like the "current" frame (the sky will still be blue, for example), it is possible to predict portions of a future frame. Since it is possible to predict what a frame might look like several frames away from the "current" frame, it is also possible to calculate what the frames "in between" should look like. For a network professional, there is a trade-off between large GOP (for example, I-B-B-P) and a small GOP (for example, I-P). A single data error anywhere in the GOP will cause a video "jump", and the smaller the GOP is more robust from a transmission point of view. But the larger GOP results in more compression, and therefore higher quality for a given data rate.

Reordering the GOP

MPEG delivers frames as they are generated. So I-Frames get generated first, then P-Frames, then B-Frames. But B-Frames "belong" in between I-Frames and P-Frames! The MPEG decoder takes care of this, and re-orders the frames. To accomplish this, it must hold on to at least one frame. Since the video is running at 30 fps, that's 0.33 msec of built-in delay.

Can You Hear Me

Audio is an integral part of MPEG, unlike MJPEG. In the previous discussion, we concentrated on video, but MPEG audio is equally important. Generally MPEG audio (so called Layer 2) is sampled at 44.1 or 48KHz, and delivered at a bit rate of 64 to 256 Kbps in mono, stereo, and several modes of "join stereo" (which saves transmission bandwidth by only encoding lower omi-directional frequencies once rather than for each channel).

A/V Sync

Video encoded using MPEG is called a "video elementary stream", and audio encoded using MPEG is called an "audio elementary stream". MPEG puts them together to create what is called a "System Stream" (MPEG-1) or a "Transport Stream" (MPEG-2) that provides a means to keep the two streams in synchronism. Each stream consists of a series of packets that are multiplexed with timestamps. These timestamps tell the MPEG decoder when to play the audio and in what order to play the video. MPEG elementary streams may also be sent via a network protocol called RTP. RTP enables synchronism between MPEG audio and MPEG video, each of which are sent separately.

Copyright VBrick Systems, Inc.

Sending It Out
With a properly constructed MPEG video stream now ready, it needs to be delivered. Depending on the needs of the network, it is delivered via Ethernet and UDP/IP, RTP, ATM, or via T1/E1, satellite or microwave. All that the payload cares about is that it gets to the other end with the least possible delay and jitter. ATM, of course, has well known QoS features, and the recent advances in DiffServ and other mechanisms makes it possible to guarantee performance network-wide.

Effects of Data Errors
In data networking, we usually don't concern ourselves with errors – we concern ourselves with *too many* errors. The data networking industry grew up in a world where we were trying to force data into noisy wires designed for telephone traffic, and as a result developed retransmission techniques to ensure data integrity. But for real-time video, retransmission is not a possibility. And for two-way and high quality video, we don't want to do a lot of receive buffering (too much delay). Thankfully, our networks have evolved and are much less prone to error. In the MPEG case, a data error will cause a lost video frame (1/30th of a second) or so and does not represent a major concern for most video applications.

Video is Not a Bandwidth Hog
A typical video stream as produced by a VBrick, will deliver, let's say 2 Mbps over a 10BaseT network via UDP/IP. This is 20% of a 10BaseT, or 2% of 100BaseT. Yes, if this is inserted into an oversubscribed hub, no one on that hub will enjoy his or her email or Internet experience. You swamped the network. But this is not true if the VBrick is on its own port on an Ethernet Switch and/or through a modern router. Ethernet switches typically cannot be oversubscribed because the sum of the ports is less than or equal to the capacity of the switch. And in the case of multicast, you already know that only those clients who request a stream receive it.

Consider how your network deals with a user on a 100 Mbps port pulling a 1G file from a server? For the duration of that transfer, the network is trying use as much bandwidth as possible, leading to huge peaks and valleys of network utilization. Video, on the other hand, is a constant, well-regulated load that can be managed.

For the most part, network professionals are concerned with WAN bandwidth usage. Indeed, if only a handful of users in a corporate network view Internet streaming video, it could exceed the capacity of a T1 access line. Video in the intranet, on the other hand, not only consumes none of this resource, but helps to relieve the WAN because the sources are local.

Video Management
Today, video network appliances appear to network managers as just another network device…fully managed via SNMP, Telnet, Web interfaces, and GUI tools. For example, a VBrick sitting anywhere on a IP network can be fully managed – not just network

Copyright VBrick Systems, Inc.

characteristics, but things like brightness, contrast, GOP, etc. A video network appliance is much like a network printer – but much less bursty.

Applications

Television, both one-way (a.k.a. "streaming") and two-way (a.k.a. "conferencing"), is being delivered over new and existing networks for a wide variety of applications. While corporate communications and training tops the list, video is widely used for security and monitoring, advertising, and entertainment.

Network professional should be honored to know that today more than ever, the future of television belongs to you.

Copyright VBrick Systems, Inc.