telchemy
Actively managing multimedia®

Application Note

▷ **Title**   **IP Video Encoding Explained**

▪ **Series** Understanding IP Video Performance

▪ **Date**   May 2008

▷ Overview

IP networks are increasingly used to deliver video services for entertainment, education, business and commercial applications.

This Application Note explains how video is encoded and transmitted over IP, describes common impairments that can affect video perceptual quality, and lists some basic approaches to IP video performance management.

▷ Contents

Introduction

IP video (a.k.a. "Video over IP") refers to any number of services in which a video signal is digitized, compressed, and transmitted over an IP network to its destination, where it can then be decompressed, decoded, and played. Typical services include IPTV, Video on Demand (VOD), streaming desktop video, and IP videoconferencing/Telepresence systems.

The ability to transmit video over IP networks opens an enormous range of possibilities for residential, commercial, and governmental applications. The quality of the video, however, is susceptible to various impairments that can occur during encoding, transmission, or decoding/playback.

This Application Note will explain the video encoding process, as well as describing some common types of impairments and performance management strategies for ensuring the quality of IP-based video.

Video Basics

Video is transmitted as a series of pictures, or frames, at a rate typically between 25-60 frames per second. For digital video, the display size is expressed as the number of horizontal and vertical pixels that comprise the screen. The higher the resolution and frame rate, the greater the bitrate required for transmission of the video.

Some video systems use interlacing to reduce the bandwidth required for transmission. Unlike progressive (noninterlaced) scanning, in which each horizontal scan

line of a video frame is drawn sequentially from top to bottom, interlaced signals are displayed by first drawing every odd scan line from top to bottom, then 'filling in' each even line, all within a fraction of a second (approximately 1/30 second for standard 480-line NTSC television broadcasts). Interlacing is used for most standard definition television as well as the HDTV 1080i broadcast format.

Video Codecs

A digital video codec (**co**der/**dec**oder, **co**mpressor/**dec**ompressor, or **co**mpression/**dec**ompression algorithm) is a device or program used to compress the video data for transmission across the IP network, and to decompress it on the receiving side so it can be played back.

A number of factors influence the choice of video codec. These include the quality requirements for the video to be delivered, the available bandwidth, and whether real-time compression is required (essential for videoconferencing or live TV broadcasts, whereas pre-recorded content such as that from a VOD server would not necessarily need to be compressed in real time).

For broadcast IP video, the most common codec standards are:

- **MPEG-2**, a widely-used standard for cable, satellite, and digital terrestrial television broadcasts, as well as for DVD/ SVCD (however, major satellite providers are in the process of transitioning to H.264/MPEG-4 AVC for their HD channel offerings). MPEG-2 is primarily used for standard resolutions, and is not optimized for low bitrates (<1Mb/second). Unlike its predecessor MPEG-1, MPEG-2 supports interlaced as well as progressive scan video.

- **H.264** (a.k.a. MPEG-4 Part 10, MPEG-4 AVC), an emerging standard designed to provide high quality video at lower bitrates—less than half the bitrate used for older standards such as MPEG-2. H.264 is the preferred standard for DSL-based IPTV service and is being adopted by satellite providers for delivery of HD programming.

- **VC-1**, a Microsoft standard supported for Windows Media Video (WMV) 9 as well as Blu-Ray Disc and HD DVD. VC-1 is not yet widely used for digital video broadcast, but is commonly used for streaming desktop video and in some Video on Demand (VOD) applications.

For videoconferencing, the following standards are commonly used:

- **H.261**, an ITU-T compression standard originally developed in 1990 for real-time encoding and delivery of video over ISDN lines. H.261 was the first standard to introduce the "macroblock," a region of 16 by 16 pixels into each video frame is divided during the compression process. Optimized for low data rates and low motion, H.261 has been replaced in most applications by newer standards such as H.263.

- **H.263**, introduced in 1995, was designed to provide the same level of image quality at half the bitrate of H.261. H.263 is widely used for medium-quality videoconferencing applications.

- **H.264** is increasingly used for modern, high-quality videoconferencing applications such as HD Telepresence.

Video Frame Types and the Group of Pictures

When video is encoded, the video frames are arranged in a series called a Group of Pictures (GOP), of a specific structure and length. A single GOP consists of an ordered sequence of one or more of the following frame types:

- **Intracoded frame (I frame)** - a frame coded independently (without reference to any other frame). The first frame in the GOP is always an I frame. I frames are the largest frame type (requiring the greatest number of bits), and can be decoded on the receiving end without requiring data from any other frame.

- **(Forward) Predictive coded frame (P frame)** - a frame coded with motion changes from the most recent I or P frame. P frames are smaller than I frames.

- **Bidirectional Predictive coded frame (B frame)** – a frame coded with motion changes from the most recent I or P frame, the following I or P frame, or a combination of both. B frames require the fewest number of bits, but quality can be affected if too many B frames are used. Unlike I and P frames, B frames are never used as reference frames for encoding other frames. (Note that this frame type is not used in H.261; H.263 employs a similar frame type called a P-B frame.)

In addition to these, the H.264 standard introduces two new frame types, "Switching I" (SI) and "Switching P" (SP), which are designed to allow the decoder to switch between video streams using different bitrates.

Each encoded video stream consists of a successive series of GOPs. GOP structure and length are variable; typical length for a single GOP is between 15-250 frames.

The following diagram depicts a typical Group of Pictures structure.
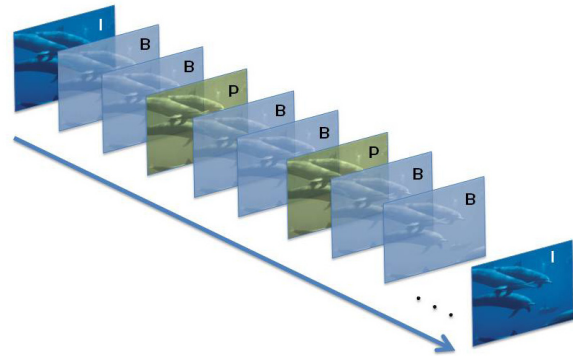

Figure 1. Typical GOP Structure

Video Compression

In order to reduce the amount of bandwidth required to transmit video over the IP network, video frames are compressed to reduce the size of the data. Much of the data comprising video frames is redundant, or "wasted space" that can be removed without a significant impact on perceptual quality.

Like digital photos, video frames contain spatial redundancy, or similarity between neighboring pixels in a single frame. They also exhibit temporal redundancy, or similarity between neighboring frames in the video sequence. The video codecs described above take advantage of both types of redundancy to compress the video in two ways—using a combination of intraframe compression (image compression) to reduce spatial redundancy in individual frames, and interframe compression (motion estimation) to reduce temporal redundancy between frames.

In the GOP, I frames are encoded using intraframe compression, while P and B frames are encoded using interframe compression.

Intraframe Compression

Intraframe compression (or encoding) is essentially the same type of compression performed on still digital images such as JPEGs. First, the video frame is divided into macroblocks that are generally 16x16 pixels in size, consisting of four 8x8 luminance blocks indicating brightness and a certain number of 8x8 chrominance blocks (typically six for MPEG) indicating color. (Intraframe compression is performed on blocks, while interframe compression—described in the following section—is performed on macroblocks.)

A Discrete Cosine Transform (DCT) is applied to each block, and the resulting coefficients are quantized. Quantization replaces many of the original coefficients with zeroes and thus allows the data to be compressed further, reducing the total number of bits that need to be transmitted and the required bitrate.

The codec at the receiving end of the video stream performs similar calculations, in reverse order, to reconstruct the video frame from the compressed data. This form of compression does result in some loss in quality in the recreated image, but in many cases the

loss is imperceptible to the human eye. (However, coarse quantization—sometimes applied to compress low bitrate video such as streaming desktop video—can result in "quantization noise" and an overall degradation of perceptual quality throughout the video sequence.)

Figure 2 illustrates the effect of quantization noise on perceptual quality. Both depict the same frame from an MPEG-2 video sequence at an original screen resolution of 720x480 at 30 frames per second. In the left-hand frame, the video sequence was encoded at a bitrate of 5 Mbits/sec. In the right-hand frame, a bitrate of 256 Kbits/sec was used, and the image shows significant distortion in the form of "blockiness."

While it is possible to use only intraframe compression (which results in a GOP structure of "IIII…"), the processing overhead and bandwidth required make this impractical for IP video. Most video codecs use a combination of intraframe and interframe compression in order to reduce bandwidth requirements while retaining the best possible image quality.

Interframe Compression

Interframe compression (or encoding) is a method for further reducing the data size and bitrate required for transmission of video over IP. In video sequences, particularly relatively low-motion scenes such
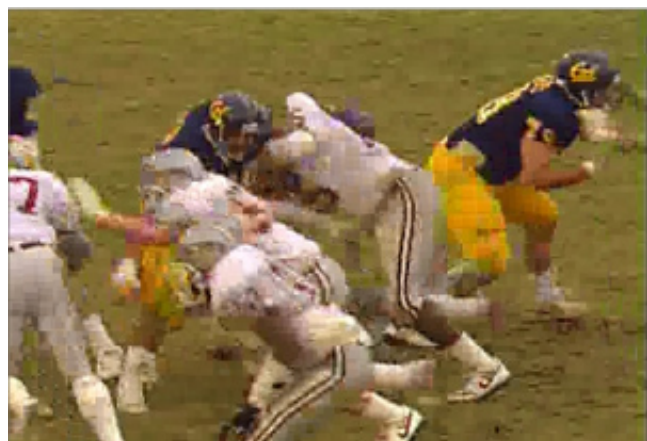

Figure 2. High bitrate (5 Mb/s) vs. low bitrate (256Kb/s) encoding

P frames are coded with motion differences from previous I or P frame

B frames are coded with motion differences from previous I or P frame, following I or P frame, or a combination of both
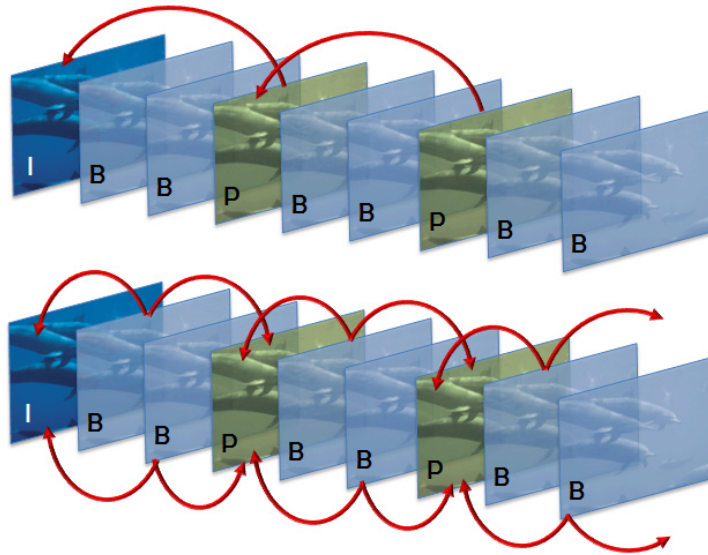
Figure 3. P and B frame coding dependencies

as footage of a news anchor, there is often very little change from one frame to the next. Rather than compressing and transmitting every frame independently, interframe compression involves encoding some frames predictively, i.e., calculating and transmitting only the required motion changes from previous (and in some cases, subsequent) reference frames.

The first frame in the GOP is always an I frame, which is encoded independently using intraframe encoding as described in the previous section. P frames are encoded with the motion differences from the previous I or P frame. B frames can be predicted bidirectionally, using information from the nearest previous I or P frame, following I or P frame, or a combination of both (see Figure 3).

The nature of interframe encoding means that encoding or transmission-related errors can have widely varying impact on the perceptual quality of the video, depending on which frame types in the GOP are affected (see Figure 5 on page 6). If an I frame is corrupted during encoding or as a result of packet loss during transmission, the error will propagate through

all remaining B and P frames in the GOP, causing distortion that may be visible for up to several seconds. An error in a P frame will propagate to any remaining B and P frames, and an error affecting a single B frame will only affect that frame (typically 15-30ms) and may not even be noticed by the viewer.

As mentioned previously, GOP structure and length can vary considerably. The use of a longer GOP can reduce bandwidth consumption, but makes the video quality more susceptible to degradation due to encoding errors or packet loss/discard, and may cause increased delay when changing channels during IPTV broadcasts. A shorter GOP reduces the impact of errors, but creates a larger number of I frames---increasing both the bandwidth requirements and the processor load on the encoder and decoder. Video performance management and testing solutions, such as Telchemy's VQmon/HD and DVQattest, can help determine the optimal GOP length for specific IP video applications.

Encapsulation and Transmission

Each compressed video frame is divided into a number of transport units, each of which is encapsulated into
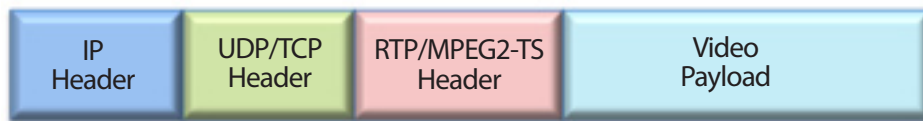
Figure 4. Typical IP Video Packet Encapsulation

a transport packet (RTP or MPEG-2 Transport, or occasionally both), then in UDP or TCP, and finally in IP, and then transmitted across the network. Figure 4 shows the encapsulation format for a typical IP video packet.

To prevent errors due to packet loss, some video systems employ Forward Error Correction (FEC), which adds redundancy to the transmission and allows some lost packets to be replaced on the receiving end.

Because the decoding of B frames sometimes requires data from "future" reference frames, frames are commonly transmitted in a different order than that of the GOP. For example, a GOP sequence of "IBBPBBPBB..." would be sent as "IPBBPBBPB..." Frames may also be temporarily stored in a frame buffer at the decoder.

Impairment Types

The quality of IP video can be affected by impairments introduced during the encoding process, during transmission of the packetized video across the IP network, and during decoding and playback of the reconstructed video signal. Some measures commonly taken to conserve bandwidth—such as the use of coarse quantization, a longer GOP, or a lower frame rate—can result in degraded quality.

This section describes some of the typical problems encountered and why they occur.

Encoding-related Impairments

Some level of distortion is inevitably introduced by the encoding/compression process; the visibility of
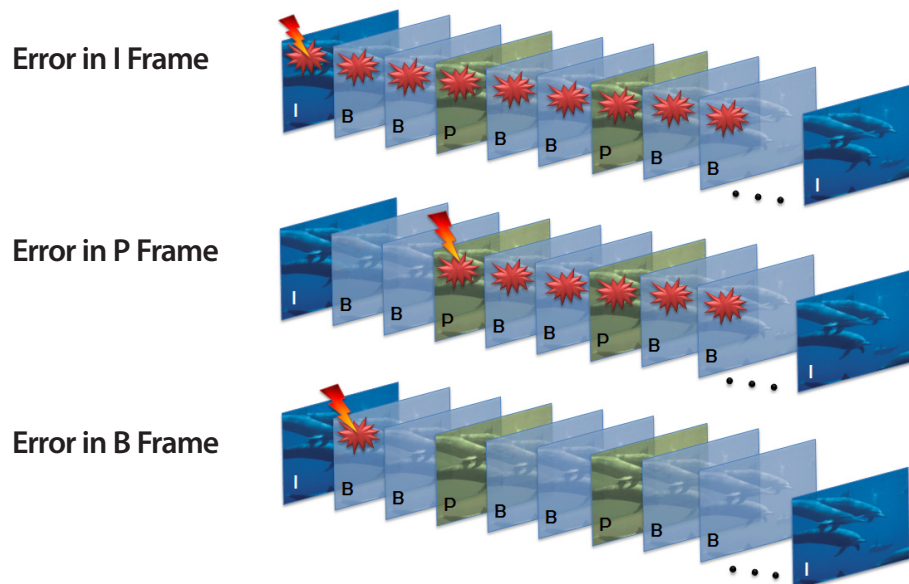

Figure 5. Error propagation through the GOP

Figure 6. Example of pixelation caused by a compression problem

distortion depends in part on the levels of on-screen detail and motion. Common problems include:

■ **Block distortion** – caused by coarse quantization, this gives the image a "blocky" appearance reflecting the underlying block structure of MPEG images.

■ **Blurring** – edges of objects appear less sharp. Commonly occurs with the use of low bitrate/frame rate algorithms and/or during high-motion video sequences.

■ **Edge busyness** – another type of distortion concentrated at the edges of objects, this is caused by quantization of the image at the boundaries between areas with a significantly different color or brightness level.

■ **Mosquito noise** – a type of edge busyness that appears as moving artifacts or noise patterns superimposed over moving objects.

■ **Ringing (Gibbs phenomenon)** – resulting from the use of DCT or Fourier transforms during the encoding process, this appears as "ripples" on either side of an edge between high contrast areas.

■ **Quantization noise** – typically occurs as visible noise or "snow" over most of the image, which may not be uniform.

■ **Jerkiness** –sequences appear as a series of "jumps" rather than displaying smooth motion. This commonly results from the use of low bitrate encoding on high motion video sequences.

■ **Pixelation** – the appearance of blocks of colored pixels or an overall "jagged" look to the image, often caused by a high compression rate or transcoding errors that occur when converting from one image format to another.

Transmission-related Impairments

Packet loss occurring during transmission can degrade the quality of the video signal in a number of ways. Frequently (but not always) caused by network congestion, lost packets can lead to missing blocks in the decoded image. If a sufficient number of blocks are missing, frame freezes or gaps in playout can occur.

As mentioned earlier, the impact of packet loss varies considerably depending on which frame types in the GOP are affected—impairments affecting I or P frames will propagate for a longer period of time and are more likely to be visible than those affecting individual B frames.

In IP video systems, some of the areas in which packet loss can occur include:

- **Home network** – IPTV service quality may suffer if the customer premises LAN has insufficient bandwidth, a home gateway with a limited internal buffer, WLAN configuration problems (or neighbors stealing bandwidth from WLAN access points), or other performance problems.

- **DSL/cable connection** – transmission problems on the access link can cause errored packets, which are typically discarded at the customer premises and appear as lost packets.

- **Ethernet Layer 1/2** – system components at the head end of the network are often connected using Ethernet. A number of Layer 1/2 problems (including duplex mismatch, faulty cables/connectors, and hardware faults) can contribute to packet loss.

- **Video bandwidth smoothing** – because I frames are much larger than P or B frames, if all video packets are sent immediately, the bandwidth peaks may overrun router buffers and lead to a higher proportion of I frame packets being dropped. The packet stream can be smoothed to provide more constant bandwidth, but this creates additional "jitter" that must be accommodated by a large enough playout buffer on the receiving end, or packets will be discarded.

Performance Management of IP Video

A good performance management system can help detect and diagnose many of the problems that affect IP video service quality, and may prevent them from occurring at all. A number of tools and technologies are available for performance management of IPTV and IP videoconferencing systems. These solutions generally fall into two main categories: passive monitoring and active testing.

Passive Monitoring

Passive, or non-intrusive, performance monitoring of live video streams can be performed using software agents embedded into endpoint equipment (such as IP set-top boxes, home gateways, and videoconferencing equipment) and/or passive probes placed at key monitoring locations in an enterprise or service provider network.

Telchemy's VQmon/HD® video performance analysis agent is one example of a passive monitoring technology that can be integrated into endpoint devices, infrastructure equipment including routers and DSLAMs, probes and analyzers, and handheld test equipment. Embedded agents monitor live video streams and report real-time video quality metrics, including Mean Opinion Scores (MOS) and other diagnostic data, without increasing network load.
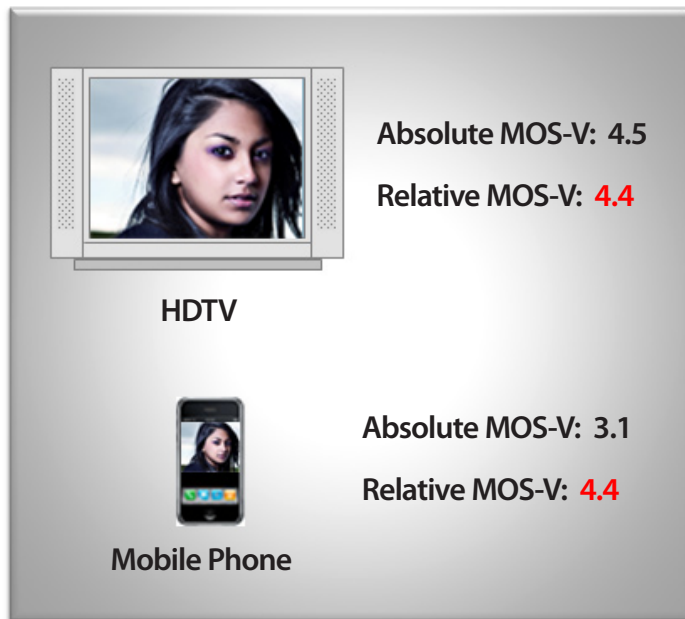
Figure 5. Absolute and Relative MOS-V

VQmon/HD's MOS scores rate video perceptual quality on a scale of 1 to 5, with 5 being "Excellent." These scores allow network managers to quickly assess the quality of the picture (MOS-V), audio (MOS-A), and combined audio-video (MOS-AV) in each video stream.

Video today is viewed in an enormous variety of formats, some with inherently higher picture quality than others—from tiny mobile handsets or portable media players to large, high-definition displays. Because higher quality displays will naturally tend to receive higher MOS scores than lower quality displays, it can be difficult to judge how well a video service is performing just by looking at the overall MOS.

VQmon/HD helps resolve this difficulty by reporting both **Absolute MOS-V** (indicating the picture quality independent of display size) and **Relative MOS-V** (which takes display size/resolution into account, indicating the picture quality relative to the ideal for that medium).

For example, a video sequence viewed on an HDTV display might receive an Absolute MOS-V of 4.5, while the same sequence viewed on a mobile phone might receive an Absolute MOS-V of 3.1 (see Figure 5). Adjusting for the display size/resolution, VQmon/HD gives both sequences a Relative MOS-V of 4.4—making it clearer that in both cases, the picture quality is close to the highest possible for that format.

Active Testing

Active test agents can be used to simulate IPTV or videoconferencing traffic in pre-deployment scenarios, for periodic service level monitoring for conformance with service level agreements (SLAs), or for advanced troubleshooting.

Telchemy's DVQattest® active test application provides powerful software test agents that can be installed on PCs or servers in an enterprise network or downloaded to residential subscriber desktops, for pre-deployment service qualification or advanced diagnosis of existing service problems.
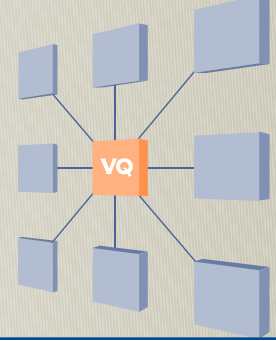
Summary

IP video presents a range of possibilities for the delivery of customized and interactive media content to subscribers, and allows businesses to minimize travel expenses by using videoconferencing/Telepresence systems for their long-distance communication needs.

Content/service providers and enterprise organizations can ensure the high quality of IP video services by implementing a robust, comprehensive performance management system—ideally, one that combines passive quality monitoring technology (such as VQmon/HD) with active test/diagnostic agents (such as DVQattest).

References

[1] ATIS-0800008 - QoS Metrics for Linear Broadcast IPTV (August 2007)

[2] ETSI TR 101 290, Digital Video Broadcasting (DVB); Measurement Guidelines for DVB Systems

About Telchemy, Incorporated

Telchemy, Incorporated is the global leader in Voice and Video over IP Performance Management with its **VQmon®** and **SQmon™** families of call quality monitoring and analysis software. Telchemy is the world's first company to provide voice and video quality management technology that considers the effects of time-varying network impairments and the perceptual effects of time-varying call quality. Founded in 1999, the company has products deployed and in use worldwide and markets its technology through leading networking, test and management product companies.

For more information, please visit www.telchemy.com.

Download application notes @ www.telchemy.com

▶ Telchemy Application Notes

Series	Understanding VoIP Performance
Title	Voice Quality Measurement
Title	Impact of Delay on VoIP Services
Title	Impact of Echo on VoIP Services
Title	Data and Fax Modem Performance on VoIP Services
Title	Voice Quality Estimation in Wireless & TDM Environments

Series	VoIP Performance Management
Title	Managing Enterprise IP Telephony
Title	Managing IP Centrex & Hosted PBX Services
Title	Managing Cable Telephony Services
Title	Managing Wireless LANs and Wi-Fi Services
Title	Six Steps to Getting Your Network Ready for Voice over IP

Series	Understanding IP Video Performance
Title	Understanding IP Video Quality Metrics
Title	TVQM Video Quality Metrics
Title	IP Video Encoding Explained

Series	IP Video Performance Management
Title	Managing IPTV Performance

▶ Information

Email
sales@telchemy.com
info@telchemy.com
Phone
+1-678-387-3000

Telchemy, Incorporated
105 Nobel Ct
Suite 100
Alpharetta, GA 30005
USA