Application Layer: Video Streaming and CDNs

Study-Ready Notes

Compiled by Andrew Photinakis

October 17, 2025

Contents

[Summary: The application layer encompasses various network services and protocols including web, email, DNS, and multimedia streaming. Video streaming and CDNs represent key modern applications addressing scalability and quality challenges.]

1 Video Streaming and CDNs: Context

- Stream video traffic: Major consumer of Internet bandwidth
  - Netflix, YouTube, Amazon Prime: 80% of residential ISP traffic (2020)
- **Challenge: Scale** - How to reach ~1 billion users?
- **Challenge: Heterogeneity**
  - Different users have different capabilities
  - Wired versus mobile connections
  - Bandwidth-rich versus bandwidth-poor users
- **Solution:** Distributed, application-level infrastructure

[Summary: Video streaming dominates Internet traffic and faces challenges of massive scale and diverse user conditions, solved through distributed infrastructure like CDNs.]

2 Multimedia: Video Fundamentals

2.1 Video Basics

- Video: Sequence of images displayed at constant rate
  - Example: 24 images/second (frames per second)
- Digital image: Array of pixels
  - Each pixel represented by bits
- Coding: Use redundancy *within* and *between* images to decrease bits needed

2.2 Video Compression Techniques

- **Spatial coding** (within image):
  - Instead of sending N values of same color (all purple), send only two values:
  - Color value (*purple*) and number of repeated values (N)
- **Temporal coding** (from one image to next):
  - Instead of sending complete frame at $i+1$, send only differences from frame i

2.3 Video Encoding Standards

- **CBR** (Constant Bit Rate): Video encoding rate fixed
- **VBR** (Variable Bit Rate): Video encoding rate changes as spatial/temporal coding changes
- Examples:
  - MPEG 1 (CD-ROM): 1.5 Mbps
  - MPEG 2 (DVD): 3-6 Mbps
  - MPEG 4 (Internet): 64 Kbps – 12 Mbps

[Mnemonic: S-T-V - Spatial (within frame), Temporal (between frames), Variable bit rate] [Summary: Video compression uses spatial (within frame) and temporal (between frames) redundancy reduction. Encoding can be constant or variable bit rate depending on content complexity.]

3 Streaming Stored Video

3.1 Basic Scenario

Figure 1: Basic video streaming architecture

3.2 Main Challenges

- Server-to-client bandwidth *varies* over time due to network congestion
  - Congestion can occur in home network, access network, network core, or video server
- Packet loss and delay due to congestion affect:
  - Playout timing (delays)
  - Video quality

3.3 Streaming Process

- Video sent from server through network
- Network delay affects arrival timing
- Video received and played out at client (e.g., 30 frames/sec)

- **Streaming:** Client plays early part while server sends later part simultaneously

[Summary: Video streaming involves simultaneous transmission and playback, facing challenges of variable network conditions that affect timing and quality.]

4 Streaming Stored Video: Technical Challenges

4.1 Continuous Playout Constraint

- During client video playout, timing must match original recording timing
- Network delays are variable (jitter)
- **Solution:** Client-side buffer to compensate for jitter

4.2 Additional Challenges

- Client interactivity:
  - Pause, fast-forward, rewind, jump through video
- Video packets may be lost and require retransmission

4.3 Playout Buffering Mechanism

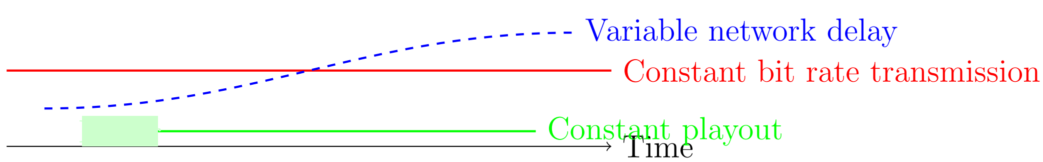

Figure 2: Playout buffering compensates for network delay variations

- Client-side buffering and playout delay compensate for:
  - Network-added delay
  - Delay jitter

[Concept Map: Network variability → Jitter → Buffering need → Continuous playout]
[Summary: Client buffers smooth out network jitter to maintain continuous playout, while supporting user controls and handling packet loss.]

5 Dynamic Adaptive Streaming over HTTP (DASH)

5.1 DASH Server Operations

- Divides video file into multiple chunks
- Each chunk encoded at multiple different rates
- Different rate encodings stored in different files
- Files replicated in various CDN nodes
- **Manifest file:** Provides URLs for different chunks

5.2 DASH Client Operations

- Periodically estimates server-to-client bandwidth
- Consults manifest, requests one chunk at a time
- Chooses maximum coding rate sustainable given current bandwidth
- Can choose different coding rates at different times
- Can request from different servers

5.3 Client Intelligence in DASH

- **When** to request chunk:
  - Prevents buffer starvation or overflow
- **What encoding rate** to request:
  - Higher quality when more bandwidth available
- **Where** to request chunk:
  - From URL server "close" to client
  - Or from server with high available bandwidth

Streaming video = Encoding + DASH + Playout buffering

[Mnemonic: C-W-W - Chunk, When, What, Where - Client decides these three aspects]

[Summary: DASH adapts video quality dynamically by breaking content into chunks encoded at multiple rates, with clients intelligently selecting optimal chunks based on current network conditions.]

6 Content Distribution Networks (CDNs)

6.1 The Scaling Challenge

- How to stream content (from millions of videos) to hundreds of thousands of *simultaneous* users?

6.2 Option 1: Single Mega-Server

- Single point of failure
- Point of network congestion
- Long (and possibly congested) paths to distant clients
- **Conclusion:** This solution doesn't scale

6.3 Option 2: CDN - Distributed Approach

- Store/serve multiple copies of videos at multiple geographically distributed sites
- Two deployment strategies:
  - **Enter Deep:** Push CDN servers deep into many access networks
    - Close to users
    - Example: Akamai - 240,000 servers in ¿120 countries (2015)
  - **Bring Home:** Smaller number (10's) of larger clusters in POPs near access networks
    - Example: Limelight

6.4 CDN Operation Process

1. CDN stores copies of content at CDN nodes
2. Subscriber requests content, service provider returns manifest
3. Using manifest, client retrieves content at highest supportable rate
4. May choose different rate or copy if network path congested

6.5 Over-the-Top (OTT) Challenges

- OTT: "Over the Top" - Internet host-host communication as service
- Key challenges coping with congested Internet from the "edge":
  - What content to place in which CDN node?

  - From which CDN node to retrieve content?
  - At which rate to retrieve content?

[Concept Map: Scaling problem → Single server fails → Distributed CDN solution → Enter deep/Bring home strategies → Manifest-based adaptive delivery] [Summary: CDNs solve scaling challenges through geographically distributed content replication, with strategies focusing on proximity to users and intelligent content delivery decisions.]

Exam Questions

Video Fundamentals

1. Compare and contrast spatial versus temporal video coding, providing examples of each.
2. What are the advantages of VBR over CBR for video encoding? When might CBR be preferred?
3. Explain how both spatial and temporal coding reduce the bitrate required for video transmission.

Streaming Challenges

1. Describe the "continuous playout constraint" and explain how client-side buffering addresses it.
2. What network factors can cause variations in video streaming quality, and how do they manifest to the end user?
3. Why is simple retransmission of lost packets problematic for real-time video streaming?

DASH and Adaptive Streaming

1. Explain the three key decisions a DASH client must make for each video chunk request.
2. How does the manifest file enable adaptive bitrate streaming in DASH?
3. What metrics might a DASH client use to determine the appropriate encoding rate to request?

CDNs and Distribution

1. Compare the "enter deep" and "bring home" CDN deployment strategies, including advantages of each.
2. Explain how a CDN helps overcome the scaling limitations of a single mega-server approach.

3. What are the main OTT challenges in content distribution, and how do CDNs address them?

[Exam Questions: Comprehensive coverage of video encoding, streaming challenges, adaptive streaming, and CDN architectures with practical application scenarios.]