# Chapter 2 Application Layer

#### A note on the use of these PowerPoint slides:

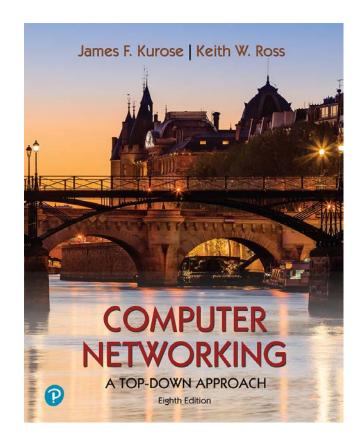
We're making these slides freely available to all (faculty, students, readers). They're in PowerPoint form so you see the animations; and can add, modify, and delete slides (including this one) and slide content to suit your needs. They obviously represent a *lot* of work on our part. In return for use, we only ask the following:

- If you use these slides (e.g., in a class) that you mention their source (after all, we'd like people to use our book!)
- If you post any slides on a www site, that you note that they are adapted from (or perhaps identical to) our slides, and note our copyright of this material.

For a revision history, see the slide note for this page.

Thanks and enjoy! JFK/KWR

All material copyright 1996-2020 J.F Kurose and K.W. Ross, All Rights Reserved



## Computer Networking: A Top-Down Approach

8<sup>th</sup> edition n Jim Kurose, Keith Ross Pearson, 2020

#### 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 ~1B users?
- challenge: heterogeneity
  - different users have different capabilities (e.g., wired versus mobile; bandwidth rich versus bandwidth poor)
- solution: distributed, application-level infrastructure







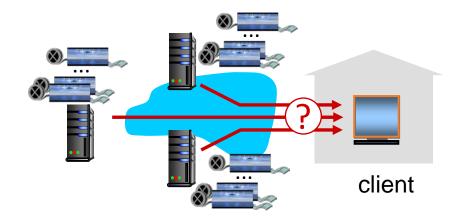


## Streaming multimedia: DASH

#### Dynamic, Adaptive Streaming over HTTP

#### server:

- 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

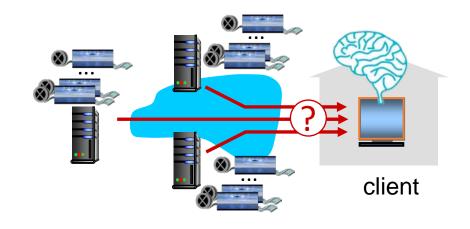


#### client:

- periodically estimates server-to-client bandwidth
- consulting manifest, requests one chunk at a time
  - chooses maximum coding rate sustainable given current bandwidth
  - can choose different coding rates at different points in time (depending on available bandwidth at time), and from different servers

#### Streaming multimedia: DASH

- "intelligence" at client: client determines
  - when to request chunk (so that buffer starvation, or overflow does not occur)
  - what encoding rate to request (higher quality when more bandwidth available)
  - where to request chunk (can request from URL server that is "close" to client or has high available bandwidth)



Streaming video = encoding + DASH + playout buffering

challenge: how to stream content (selected from millions of videos) to hundreds of thousands of simultaneous users?

- option 1: single, large "mega-server"
  - single point of failure
  - point of network congestion
  - long (and possibly congested) path to distant clients

....quite simply: this solution doesn't scale

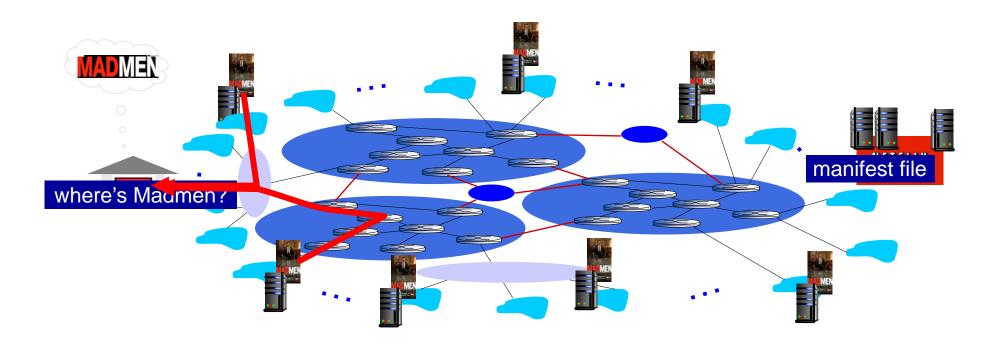
challenge: how to stream content (selected from millions of videos) to hundreds of thousands of simultaneous users?

- option 2: store/serve multiple copies of videos at multiple geographically distributed sites (CDN)
  - enter deep: push CDN servers deep into many access networks
    - close to users
    - Akamai: 240,000 servers deployed in > 120 countries (2015)
  - *bring home:* smaller number (10's) of larger clusters in POPs near access nets
    - used by Limelight





- CDN: stores copies of content (e.g. MADMEN) at CDN nodes
- subscriber requests content, service provider returns manifest
  - using manifest, client retrieves content at highest supportable rate
  - may choose different rate or copy if network path congested

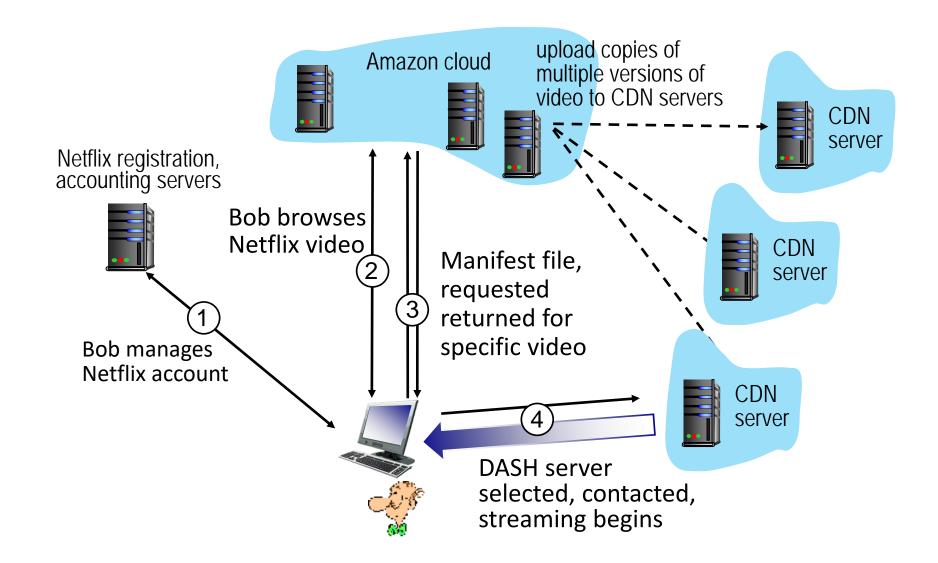




OTT challenges: coping with a congested Internet from the "edge"

- what content to place in which CDN node?
- from which CDN node to retrieve content? At which rate?

## Case study: Netflix



#### CDN content access: a closer look

Bob (client) requests video http://netcinema.com/6Y7B23V

video stored in CDN at http://KingCDN.com/NetC6y&B23V

