



A Case for a Coordinated Internet Video Control Plane

Xi Liu, Florin Dobrian, Henry Milner: Conviva

Junchen Jiang: CMU

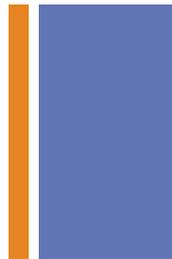
Vyas Sekar: Intel Labs

Ion Stoica: Conviva, UC Berkeley

Hui Zhang: Conviva, CMU



A Different Service Requirement



- Traditional workloads focus either on latency (e.g., interactive sessions or short web transfers) or on transfer completion time (e.g., long file transfers)
 - application data units are large enough to amortize latency effects
 - completion time does not capture rebuffering-induced interruptions
- Streaming video introduces new metrics, and requires the quality to be sustainable
- Mismatch between special streaming infrastructure and HTTP

+ Our Observations

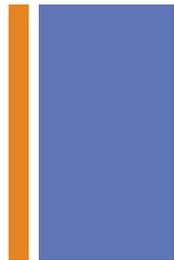


- Significant spatial diversity in CDN performance and availability across different geographical regions and ISPs,
- Substantial temporal variability in the CDN performance and client-side network performance
- Poor system response to overload scenarios when there are “hotspots” of client arrivals in particular regions or ISPs.

*Purely server- or client-driven selection and adaptation are unlikely to be sufficient – **We need a control plane with global view***



Questions We Ask



- 1. What parameters can we adapt; e.g., bitrate, CDN?
- 2. When are these parameters optimized; e.g., at video startup or midstream?
- 3. Who chooses these parameters; e.g., client or server?

+ Contributions of this work



- Measurements to expose the shortcomings of today's video delivery infrastructure that motivate the need for a video control plane.
- Using an extrapolation approach to establish the potential room for improvement.
- Corroborating these potential gains under a concrete (but simplified) operation model.

+ What we observe

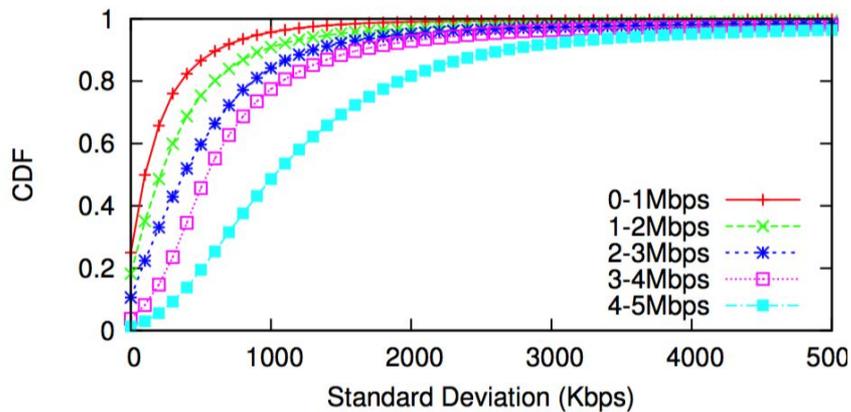


- 40% of the views experience at least 1% rebuffering ratio, and 20% experience at least 10% rebuffering ratio.
- 23% of the views wait more than 5 seconds before video starts, and 14% wait more than 10 seconds.
- 28% of the views have average bitrate less than 500Kbps, and 74.1% have average bitrate less than 1Mbps.

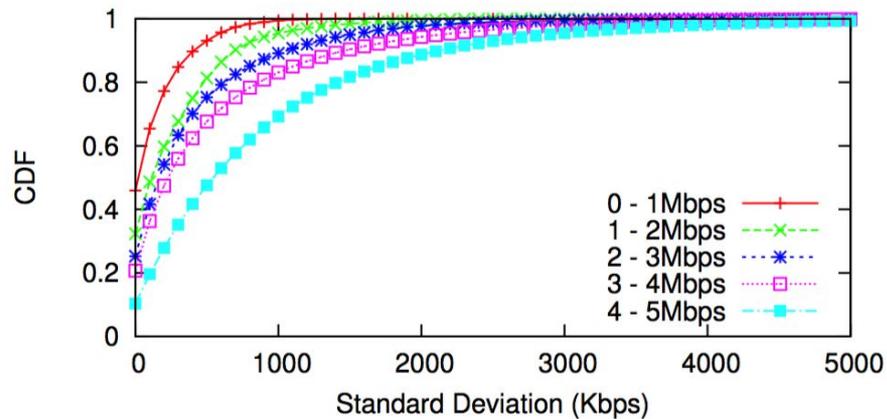
+ Source of Quality Issues



- Need for both inter & intra bitrate adaption



(a) Intra-session

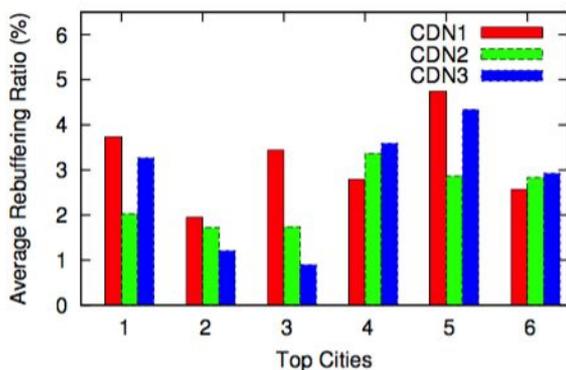


(b) Inter-session

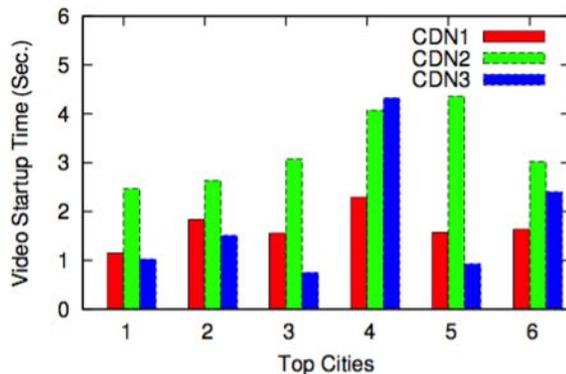
+ Source of Quality Issues



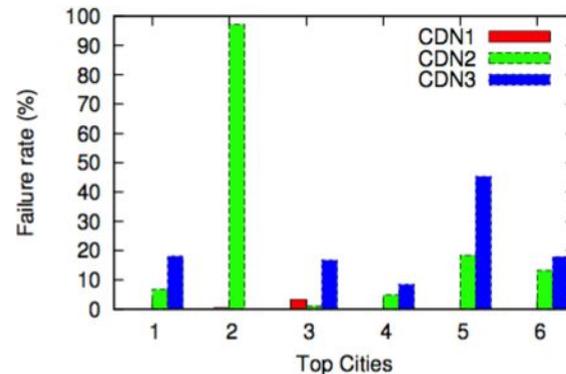
■ CDN variability over space



(a) Rebuffering Ratio



(b) Video Startup Time



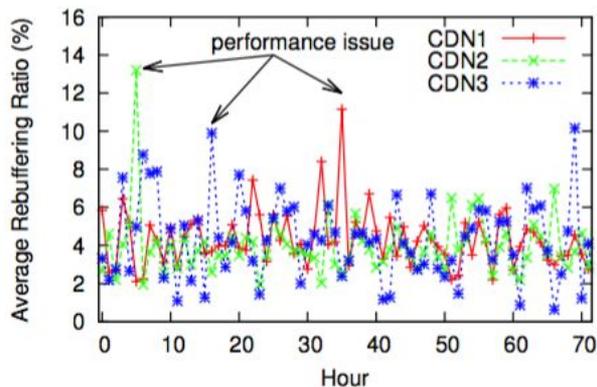
(c) Video Start Failure Rate

Figure 3: CDN performance can vary substantially across different geographical regions

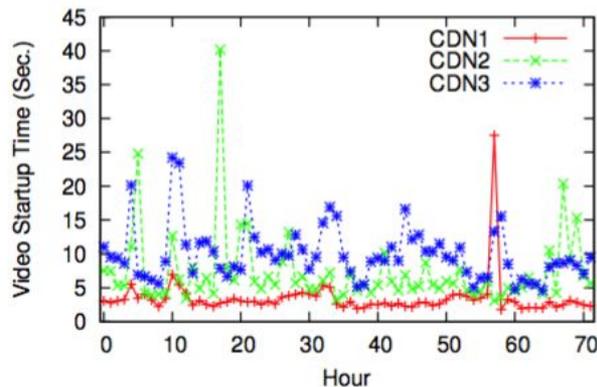
+ Source of Quality Issues



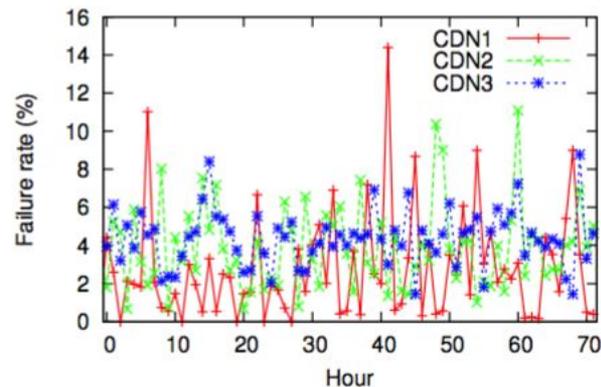
■ CDN variability over time



(a) Rebuffering Ratio



(b) Video Startup Time



(c) Video Start Failure Rate

Figure 4: CDN performance within a given geographical region can vary significantly over time as well

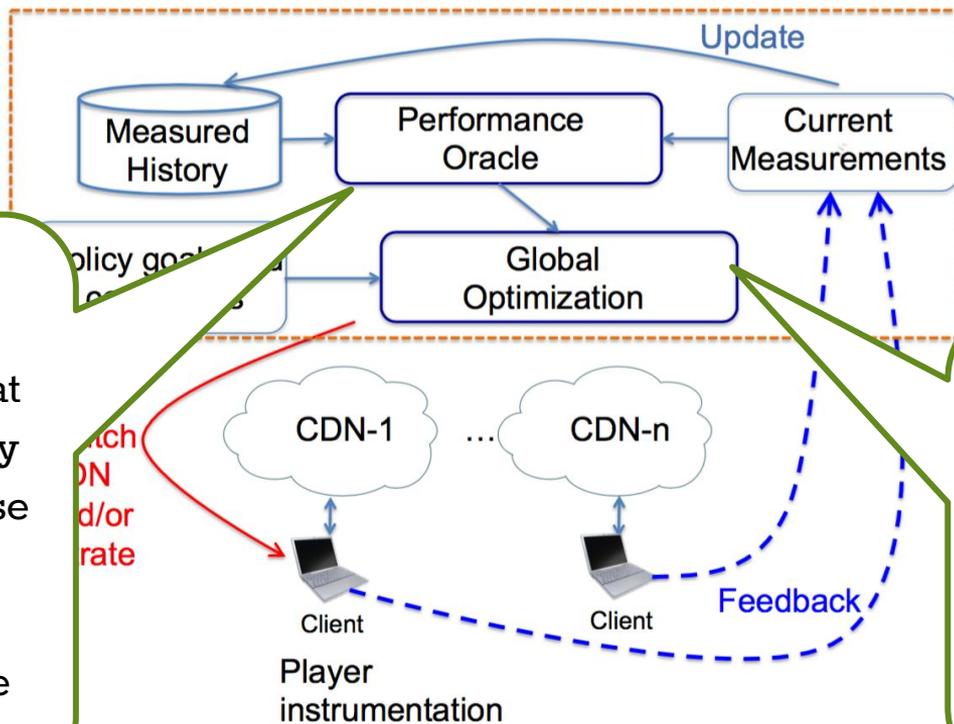
+ Design Space

- What:
 - Choice of bitrate
 - Choice of CDN(s) to serve
- Where:
 - Startup phase
 - Midstream
- Who:
 - Client
 - Control plane

What parameter?	Who chooses?	When to choose?
CDN, Bitrate	Client	Startup
CDN	Client	Startup
Bitrate	Client	Midstream
CDN, Bitrate	Control Plane	Startup
CDN	Control Plane	Startup
Bitrate	Client	Midstream
CDN	Control Plane	Midstream
Bitrate	Client	Midstream
CDN, Bitrate	Control Plane	Midstream

Table 3: Some examples from the overall design space for optimizing video delivery quality. We do not consider the cases where the client chooses the CDN and the control plane chooses the bitrate.

+ A Control Plane



predict the performance that a given user may observe if it chose a different combination of CDN and bitrate

assign each user a suitable CDN and bitrate that maximizes some notion of global utility for the content providers and consumers,

Figure 6: Overview of a video control plane

+ Approach: Estimation & Extrapolation

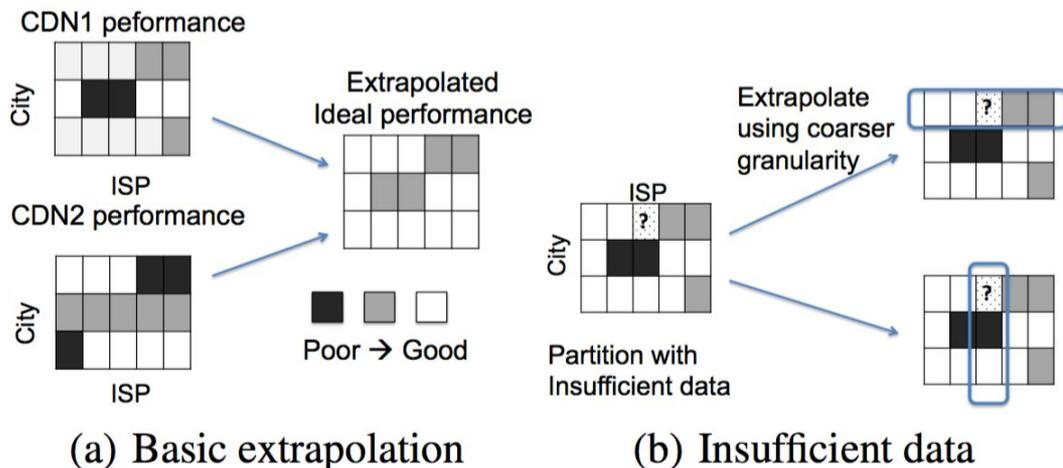
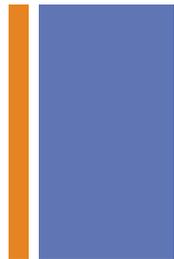


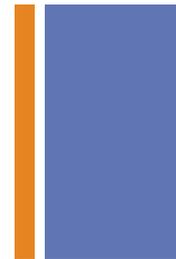
Figure 7: There are two user attributes: ISP and city. For each combination of attribute values, we want to analyze the potential improvement by choosing a better CDN (a). If a combination does not have sufficient data, we identify a suitable coarser level (b). In the simplest case, the extrapolation uses the mean of the distribution within each partition.

+ Performance Estimation



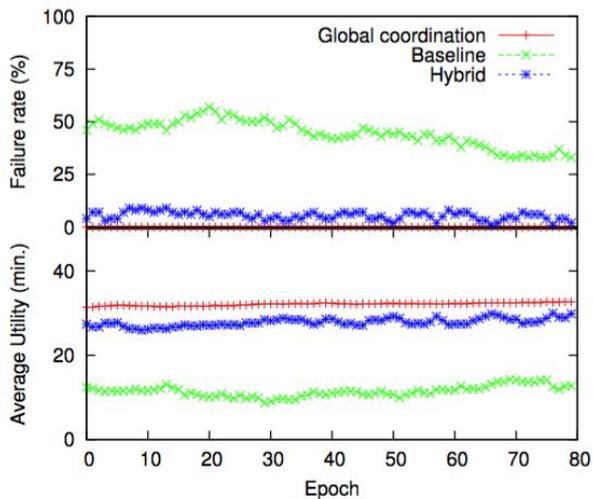
- Current performance is not visible
 - extending the hierarchical approach from the previous section to make predictions based on recent historical measurements of specific CDN-client-bitrate combinations
- CDN load
 - CDN performance shows a roughly thresholded behavior: learn threshold from history
- Bitrate not considered
 - bitrate as an additional attribute to our decision process

+ Global Optimization

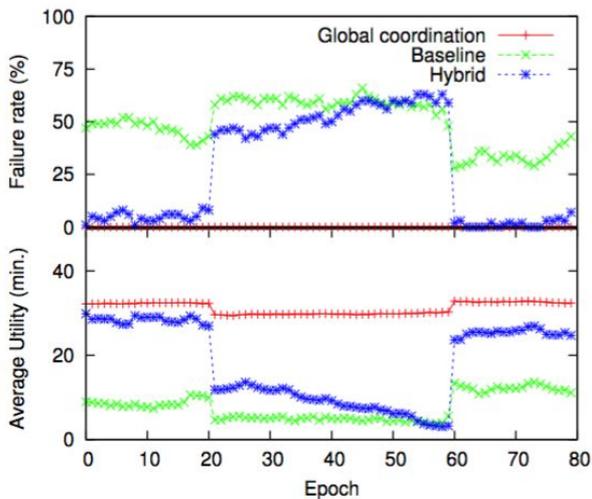


- Fairness (i.e., do not deny clients if there is sufficient capacity)
- Efficiency (i.e., maximize aggregate utility)
 - $Utility = -3.7 * BufferRatio + BitRate / 20$
 - Pick the combination of client and CDN/bitrate setting that provides the largest incremental contribution to the global utility function

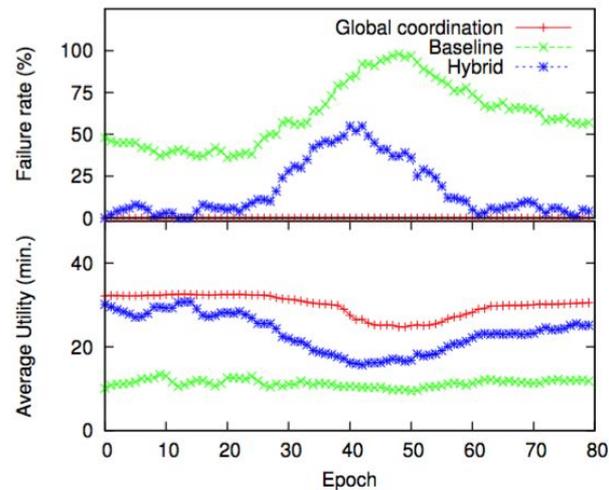
+ Simulation Results



(a) Common case



(b) CDN degradation

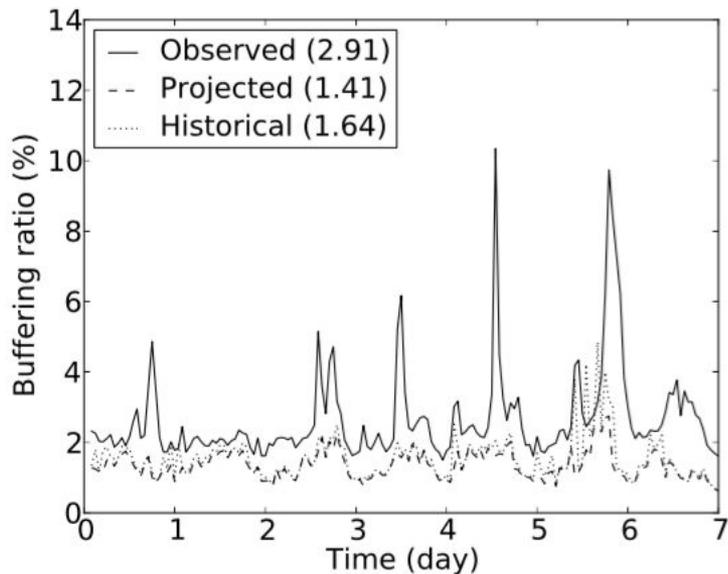


(c) Flash crowd

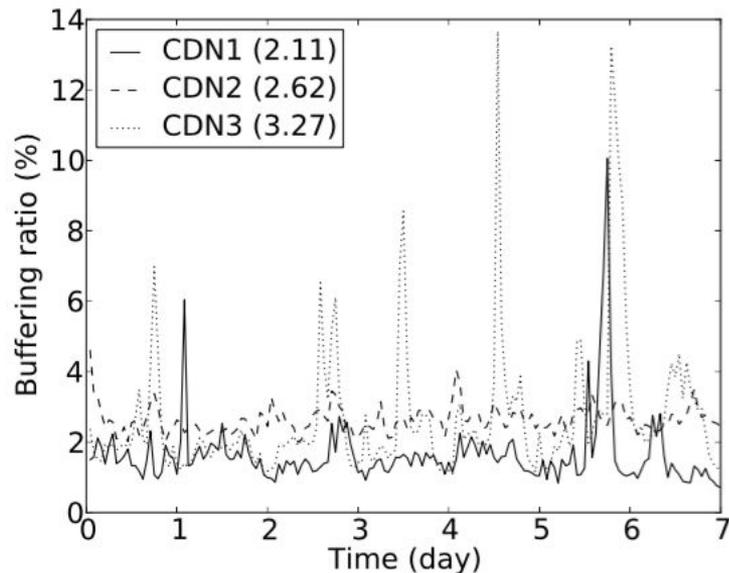
Figure 9: Simulation results of three scenarios. For each scenario, we show the performance of baseline, hybrid and global coordination in terms of failure ratio and average utility value. (The utility metric is in units of expected minutes of playing time.)



Simulation Results



(a) Prediction vs. Optimal



(b) Performance of different CDNs

+ Discussion



- Scalability
 - Unoptimized implementation in Java takes \uparrow 30s to run the global optimization for 10, 000 clients, 4 CDNs, and 5 bitrates
- Switching Tolerance
 - No good quantitative understanding on the tradeoff so far
- CDN interactions
 - No single CDN can provide the required cross-CDN optimization