Improving Video Quality by Information Sharing

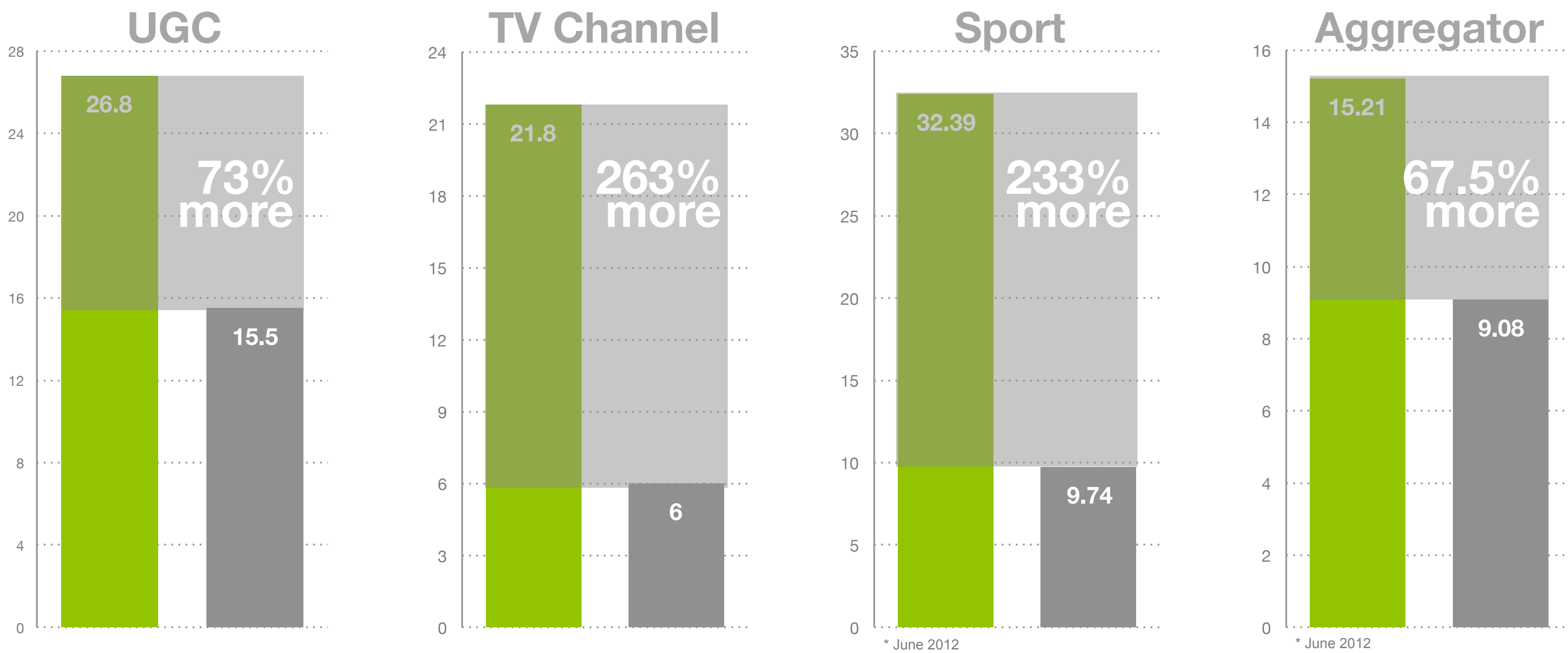
Ion Stoica, CTO
(also at UC Berkeley and Databricks)

Across many sites

We’ve seen 3 patterns

While monitoring more than 4 Billion streams per month

1 Viewers **watch more** when video is **not interrupted**



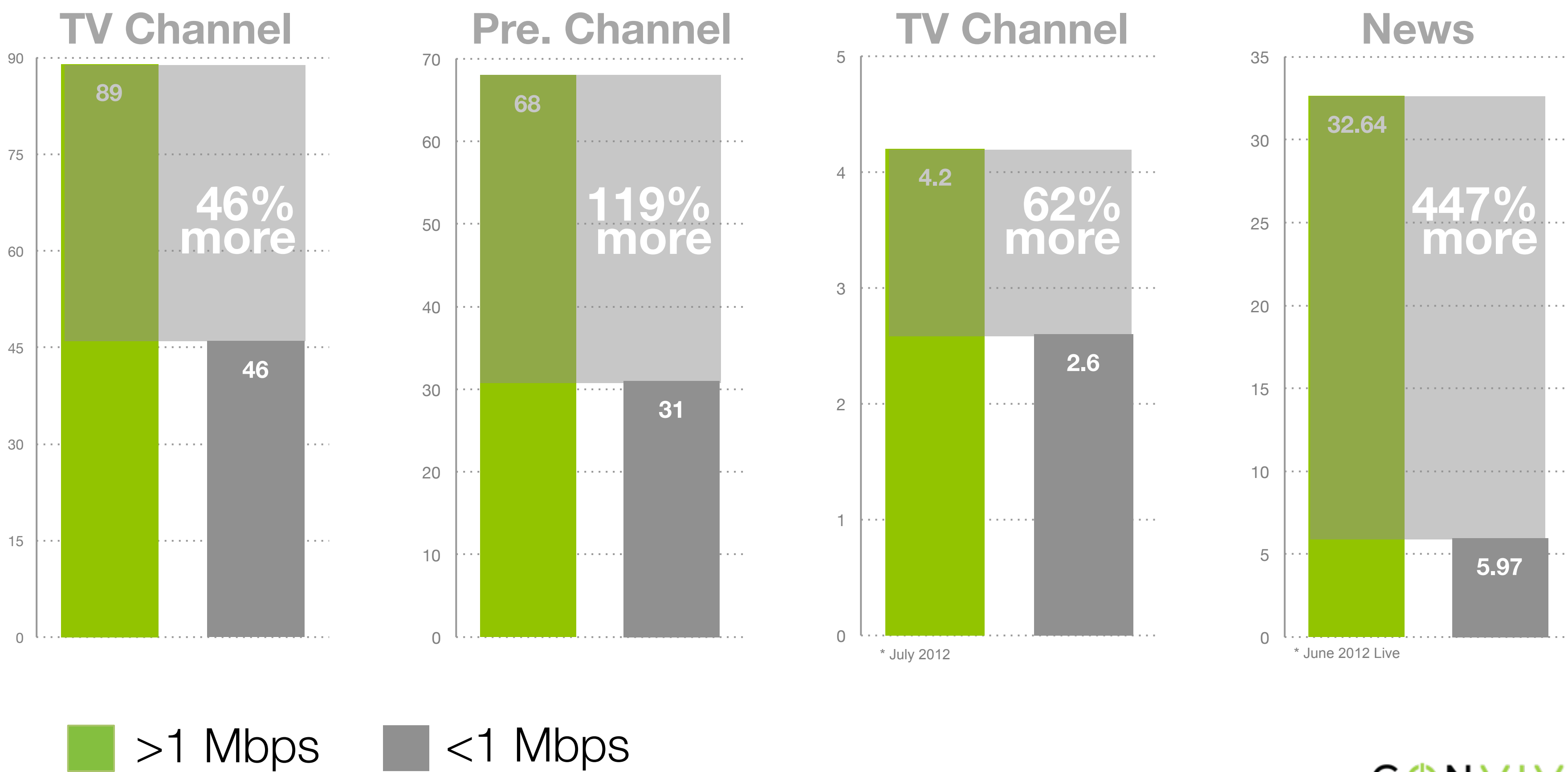
Buffering Impacted Views (BIV): > 2% buffering or > 5s continuous buffering

non-BIV BIV

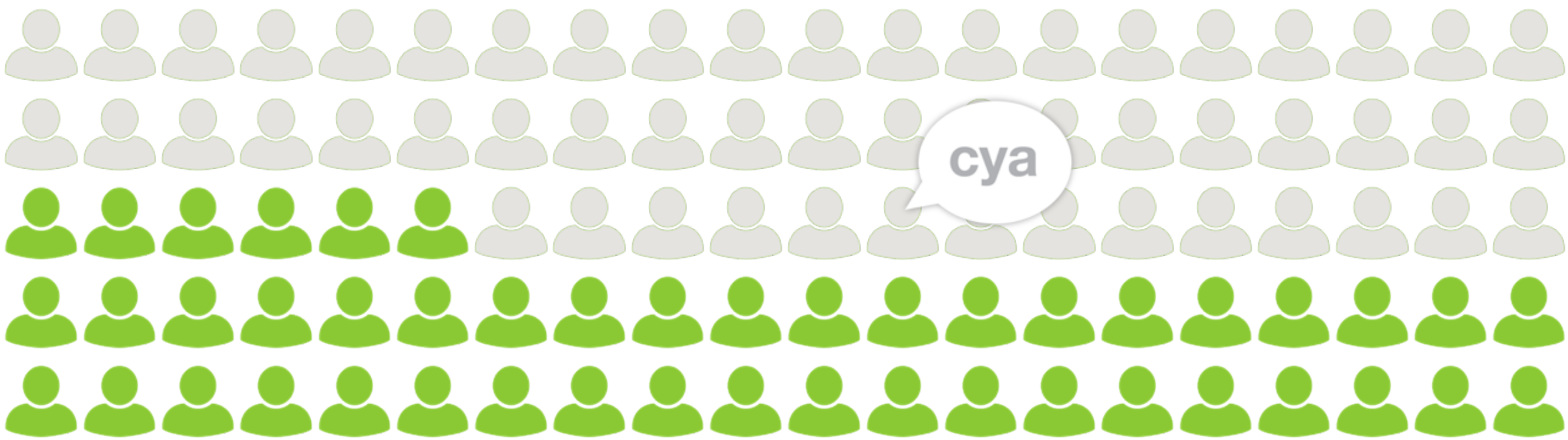
CONVIVA

2

Viewers **watch more** when video is **higher definition**



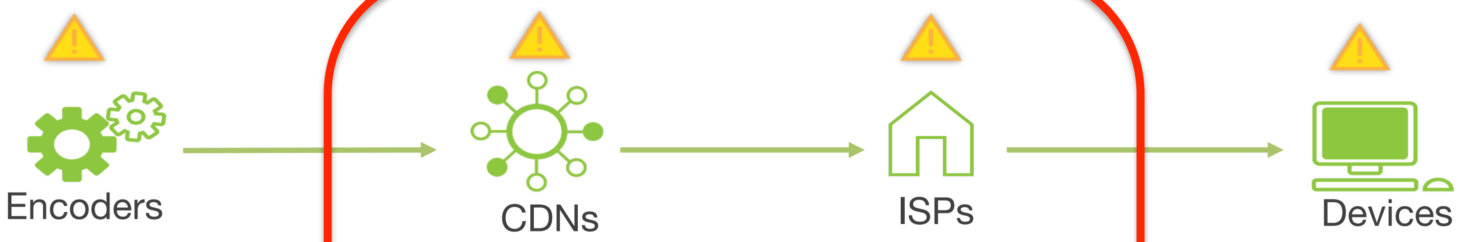
3 Viewers leave and do not return to sites when **video fails to start**



Viewers who experience a single video start-up failure
return 54% less

CONVIVA

Every step of the delivery chain presents unique challenges to delivering video **with low interruptions and high bit-rate**



Inaccurate choice of bit-rates & poor matching of screen sizes & resolution to bit-rates

Wide **variation** in CDN performance across various ISPs with peak load, type of traffic and time of day

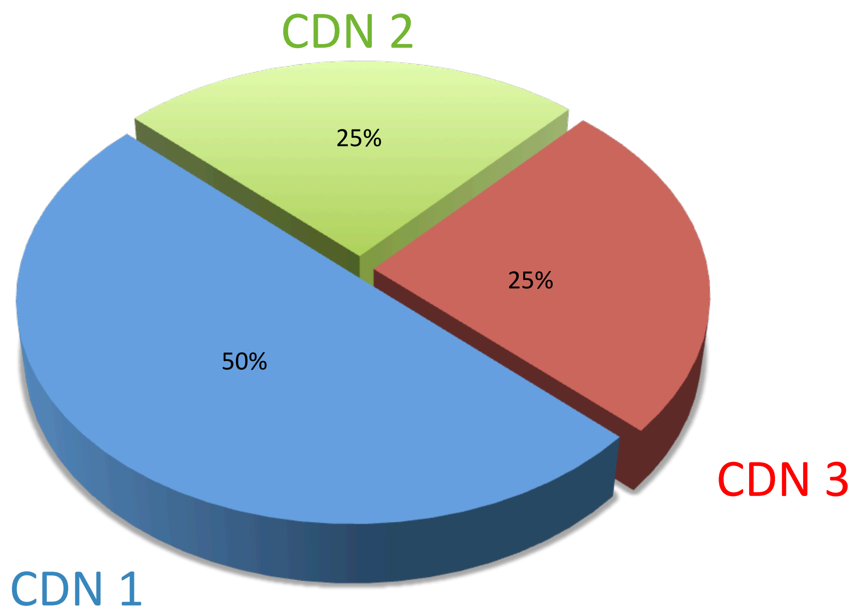
Great **diversity** in access link speeds (fiber, cable, dsl, 3G/Wi-fi, corporate networks …)

Heterogeneity of screen sizes and rendering performance across PCs, mobile hand-sets, tablets, game consoles, set top boxes, connected TVs …

CONVIVA

CDNs Vary in Performance over Geographies and Time

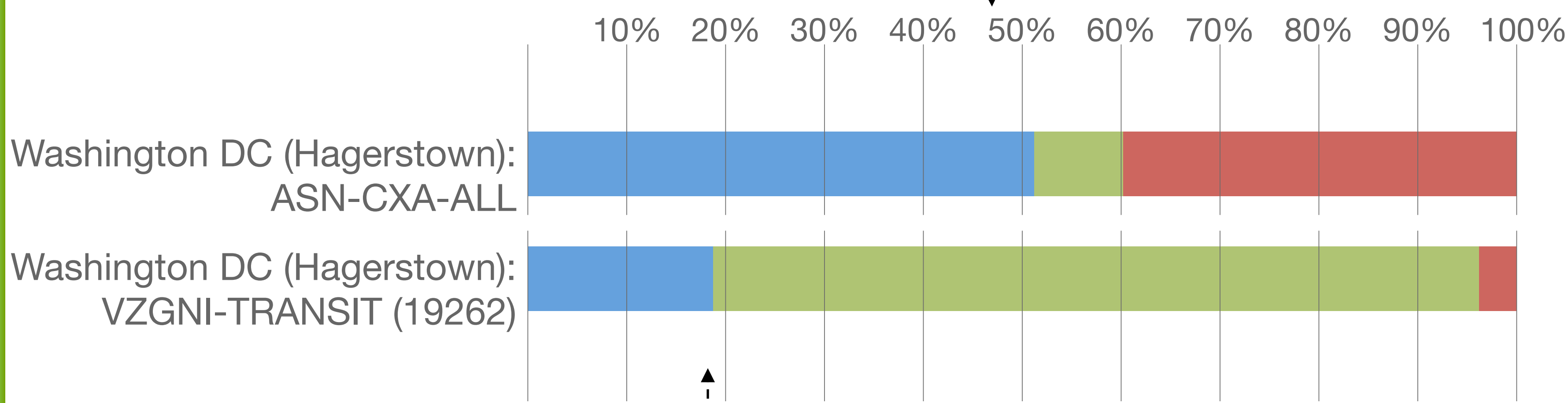
- Metric: buffering ratio
- One month aggregated data-set (2011)
  - Multiple Flash (RTMP) customers
  - Three major CDNs
- 31,744 DMA-ASN-hour with > 100 streams from each CDN
  - DMA: Designated Market Area
- Percentage of DMA-ASN-hour partitions a CDN has lowest buffering ratio



There is no single best CDN across geographies, network, and time

Washington, DC viewer experience differed greatly…

Comcast viewers got the best streams from **CDN 1** 51% of the time and only 9% from **CDN 2**

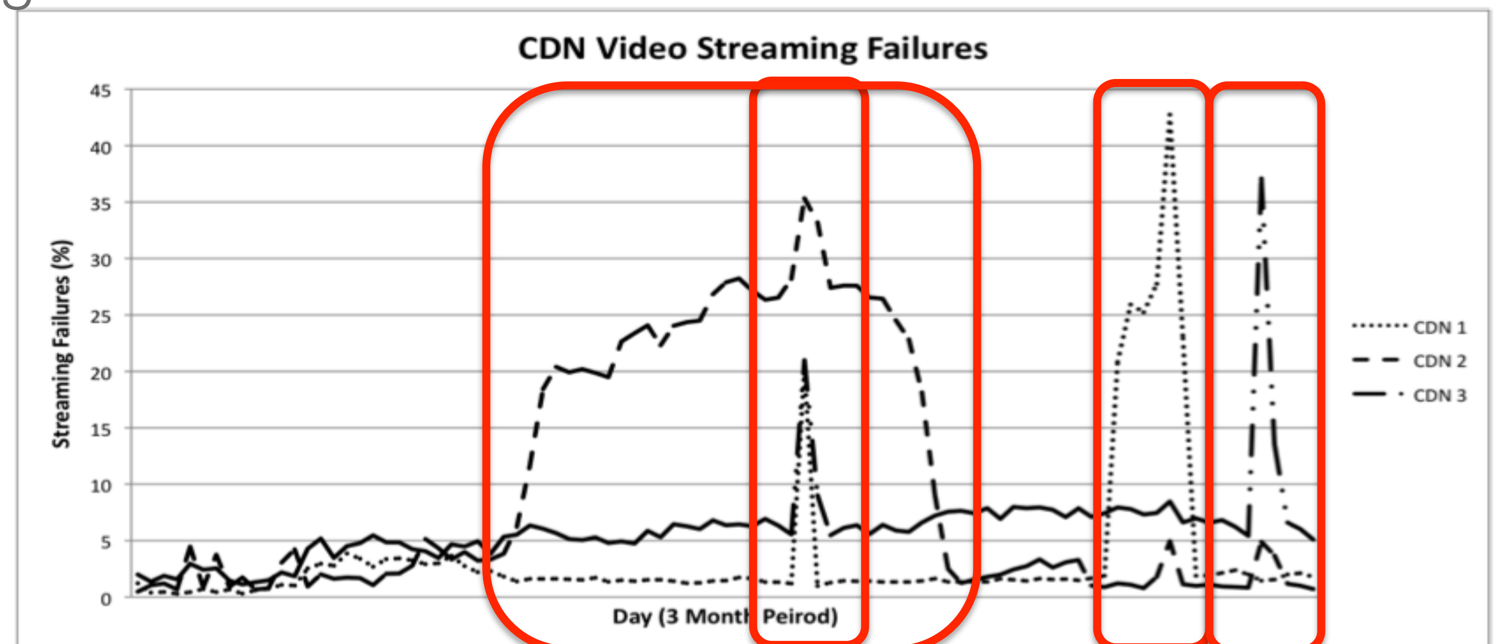


Verizon users got the best streams from **CDN 1** only 17% of the time and 77% from **CDN 2**

There is no single best CDN in the same geographic region or over time

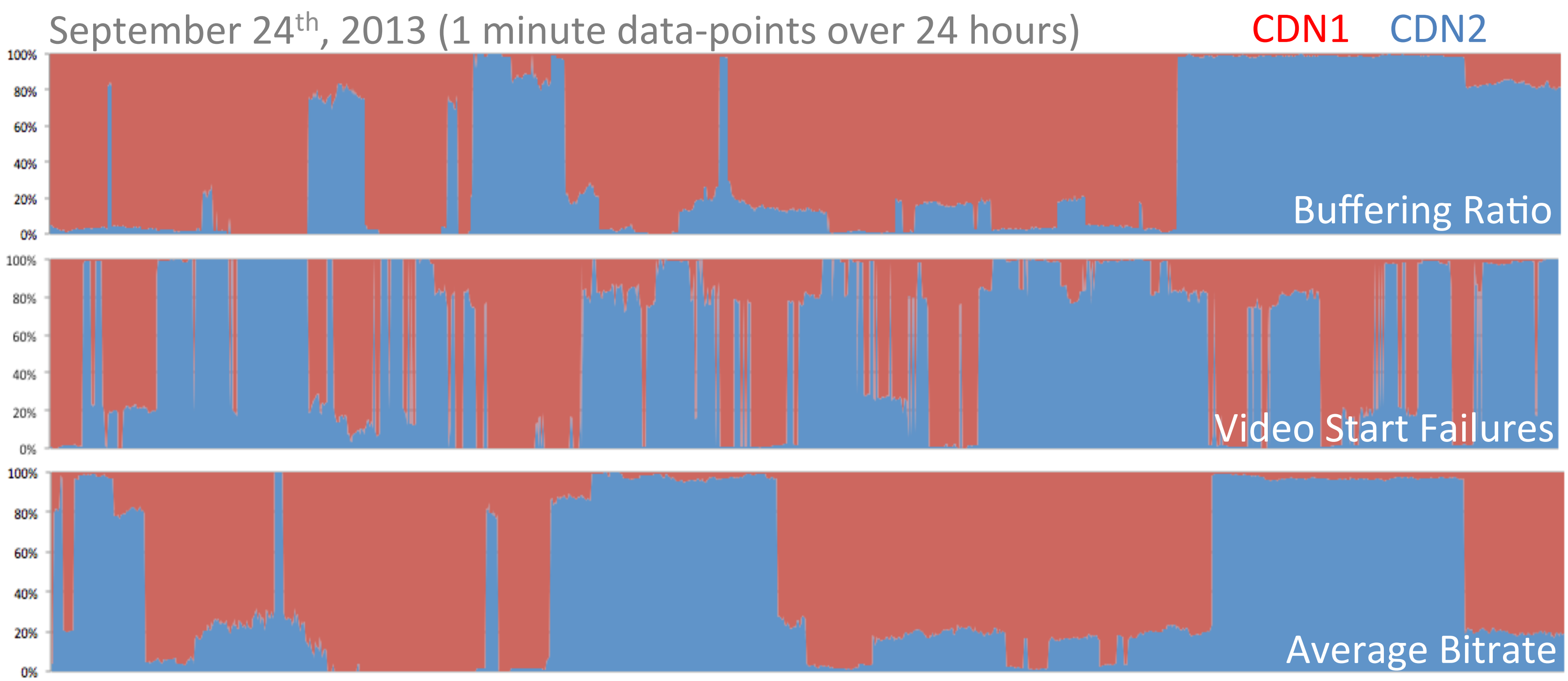
CDN Streaming Failures Are Common Events

- % of stream failures: % of streams with **video start failures**
- Three months dataset (May-July, 2011) for a premium customer using Flash



CDN (relative) performance varies greatly over time

Performance Changes Minute-by-minute



- For the same ISP, CDN performance varies minute by

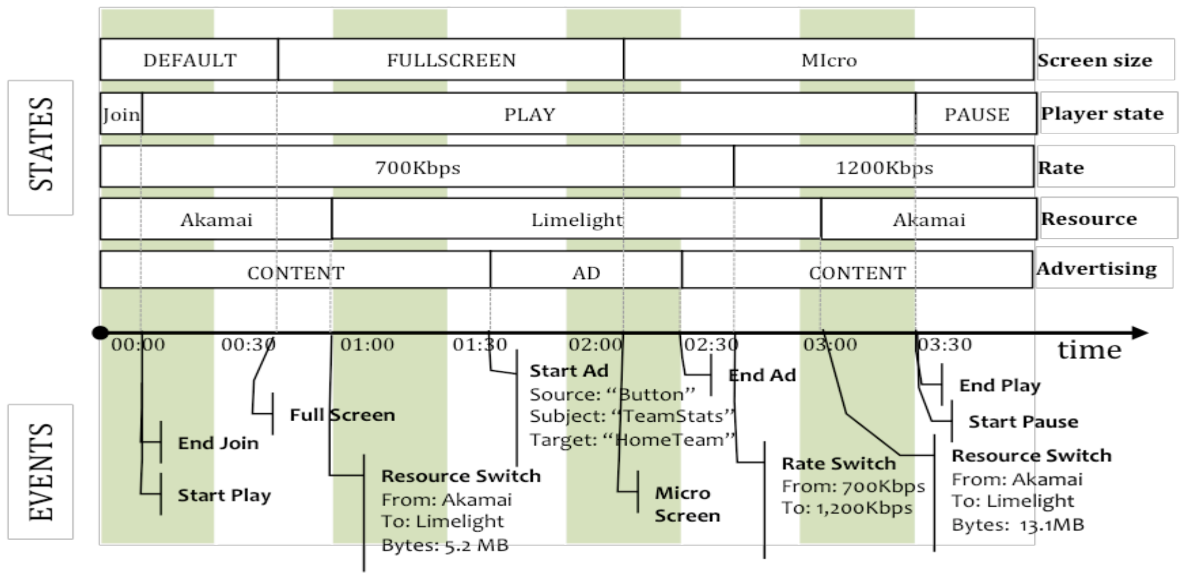
Different quality metrics no always correlated

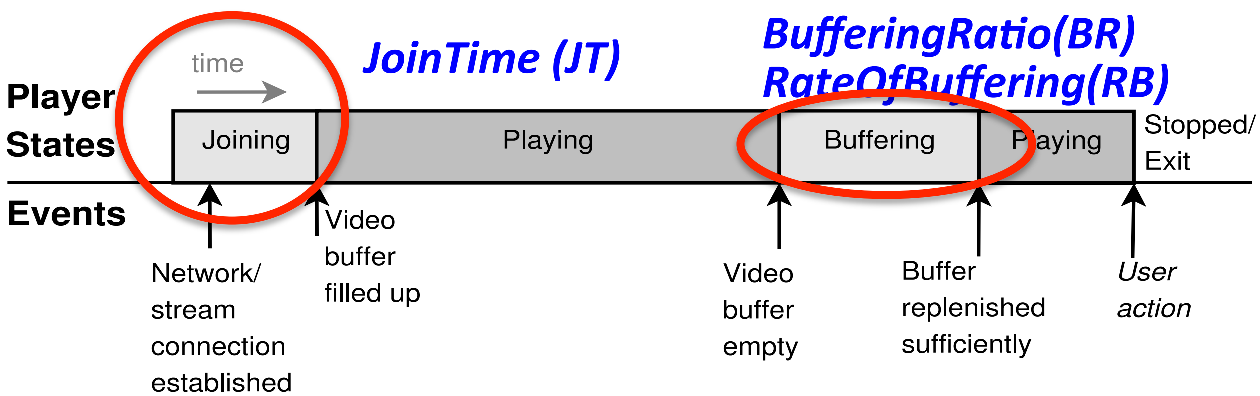
Conviva Approach to Optimize Viewer Experience

Viewer Centric Position

- True end-point sensor to see what the viewer sees and inform the source in real time

Real-time Measurement from Every Viewer





CONVIVA

Conviva Approach to Optimize Viewer Experience

Viewer Centric Position

- True end-point sensor to see what the viewer sees and inform the source in real time

Continuous Optimization

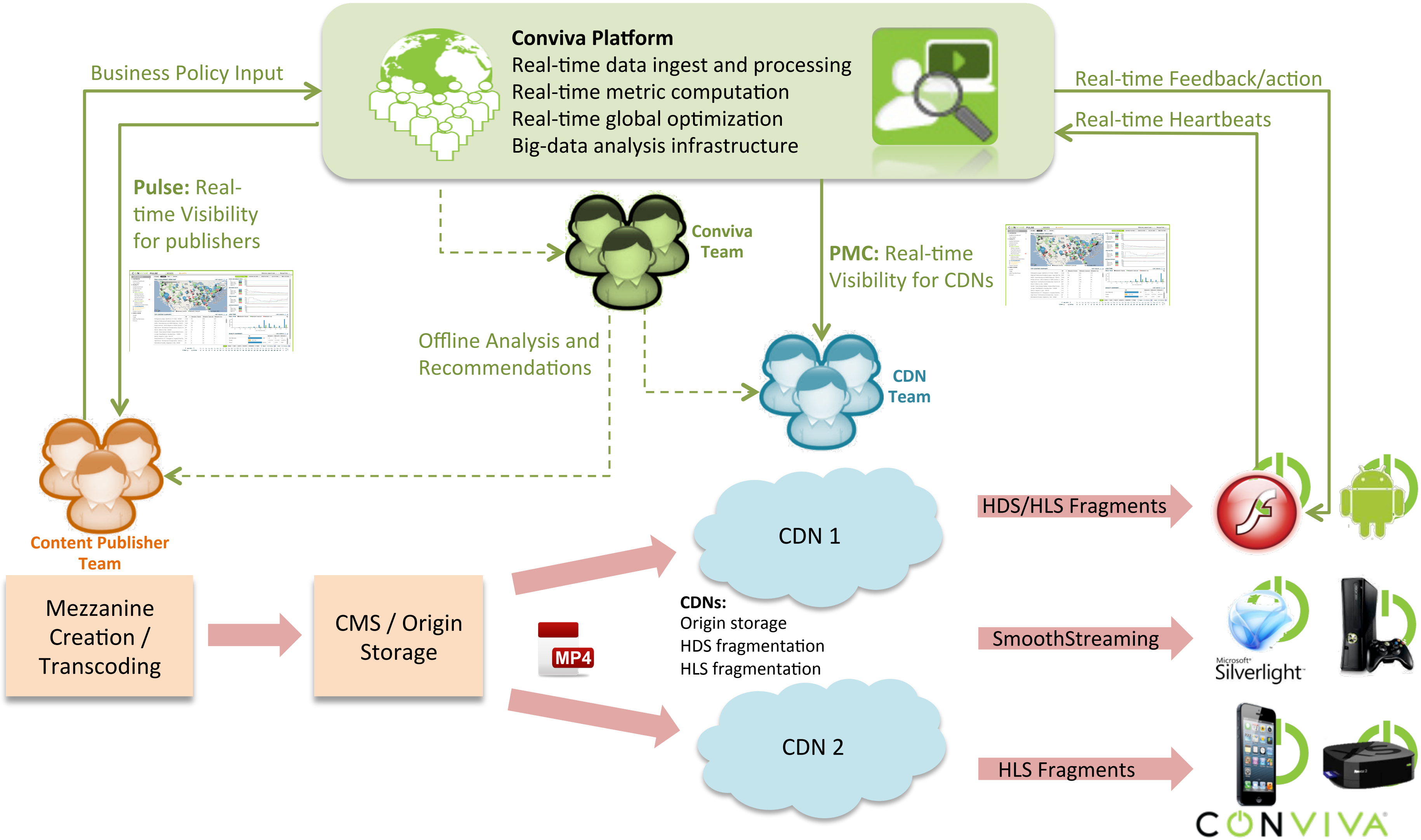
- Adjustments to streams every second to account for local environment and Internet variables

Global View and Policy Control

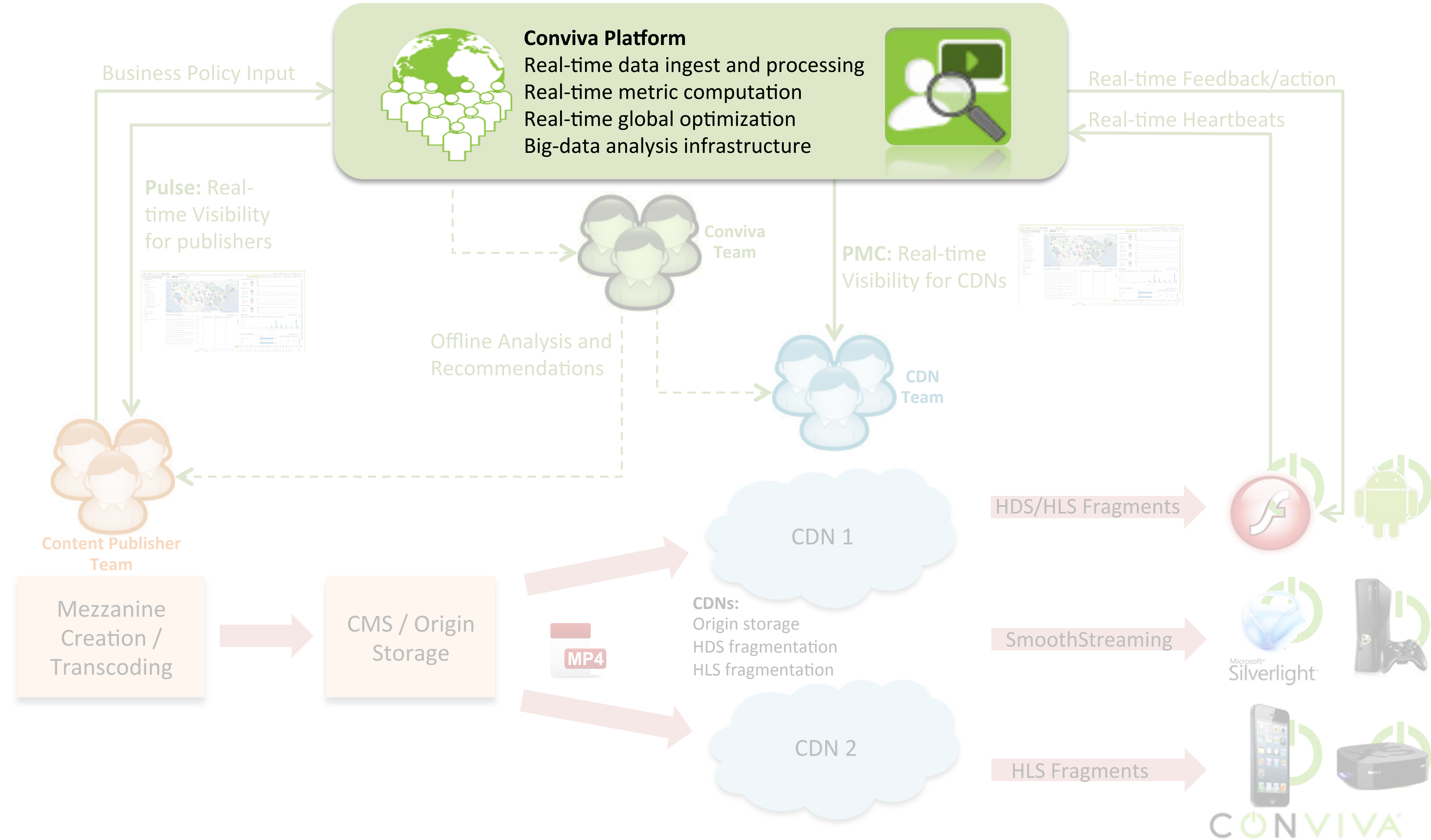
- Dynamic policy control based on real-time patterns across viewers, affiliates, and networks

CONVIVA

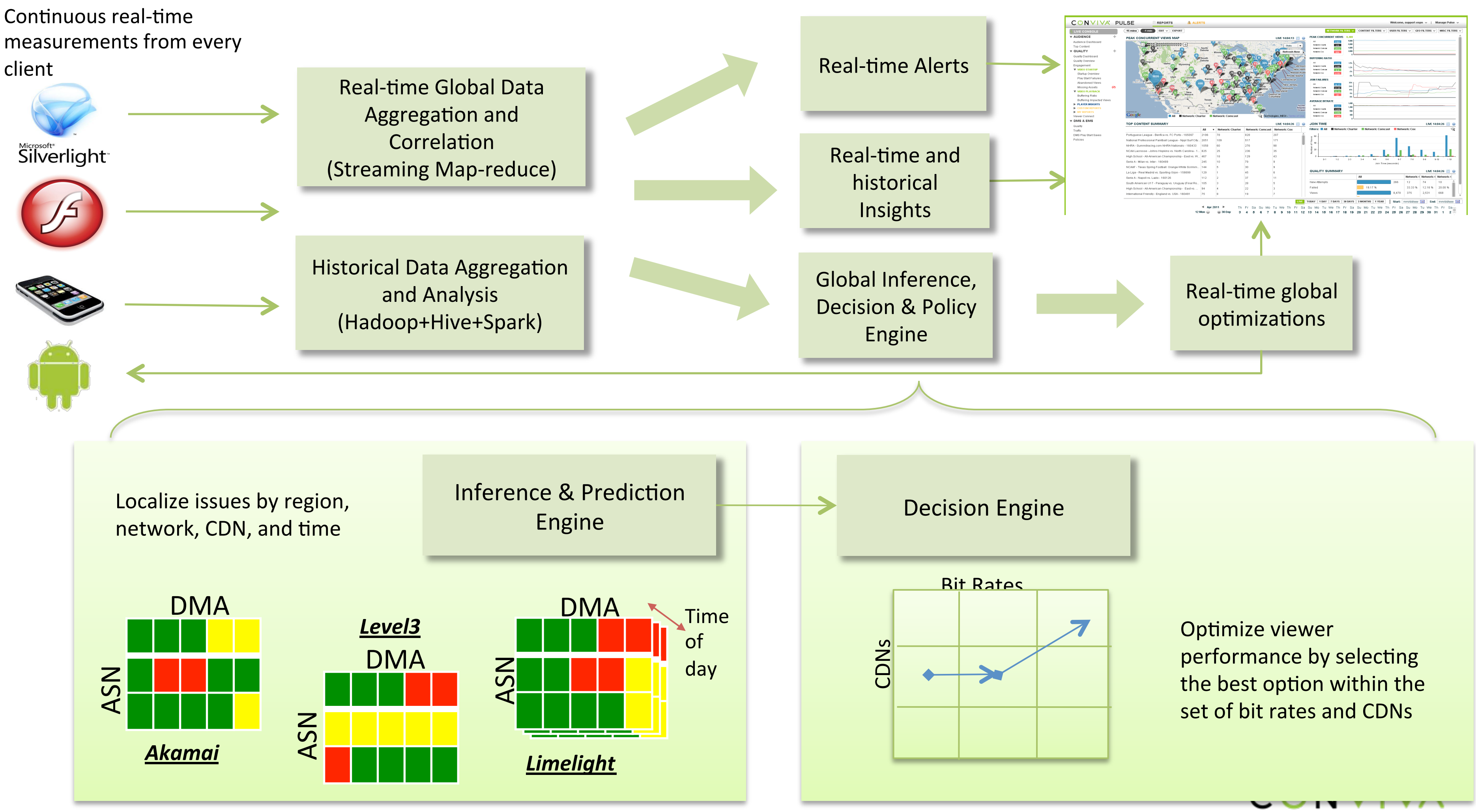
IP Video Streaming Architecture



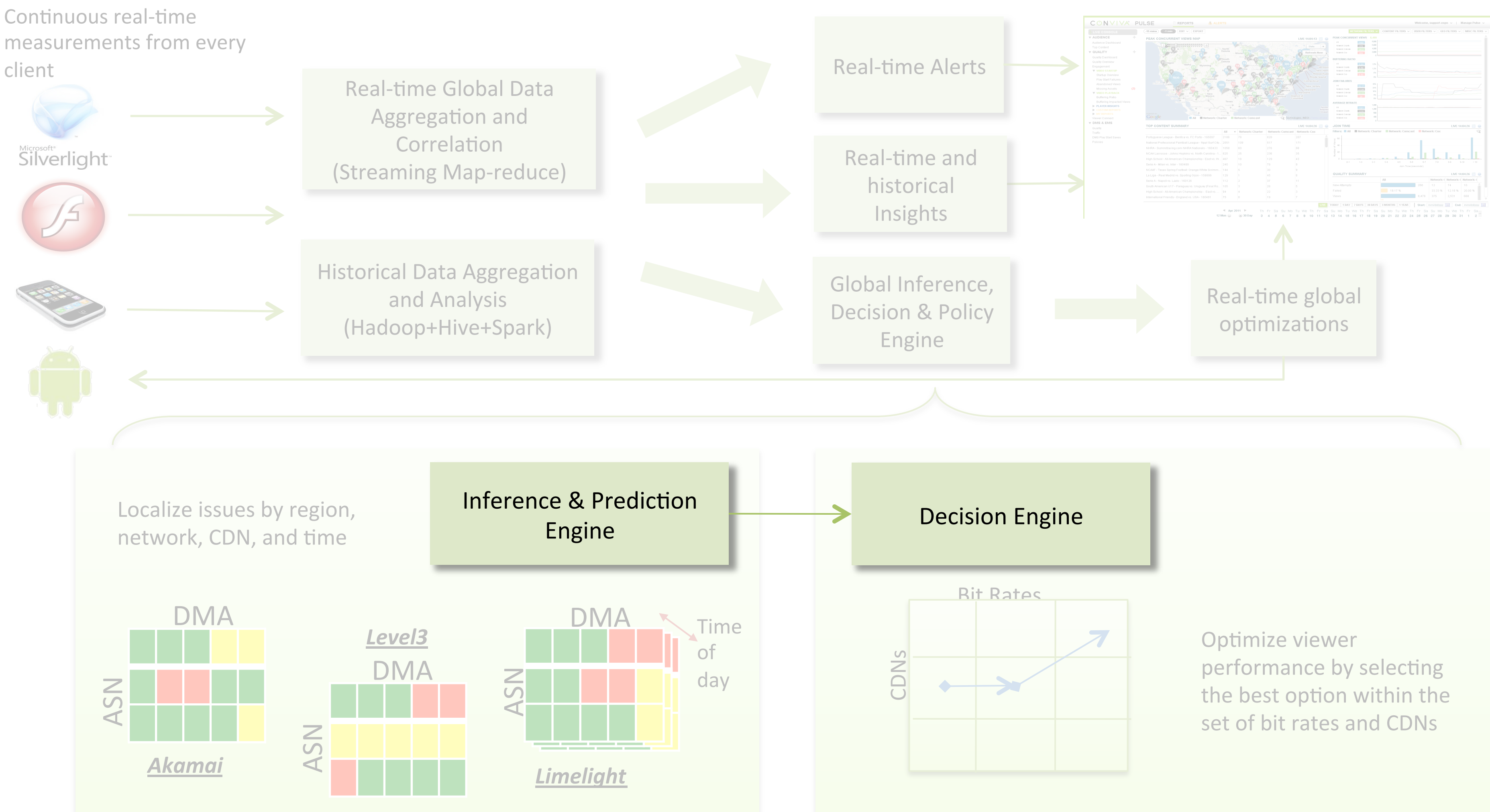
IP Video Streaming Architecture



Conviva Platform



Conviva Platform



Key Idea behind Inference & Prediction Engine

Share quality information across views

Use quality information from existing views to predict

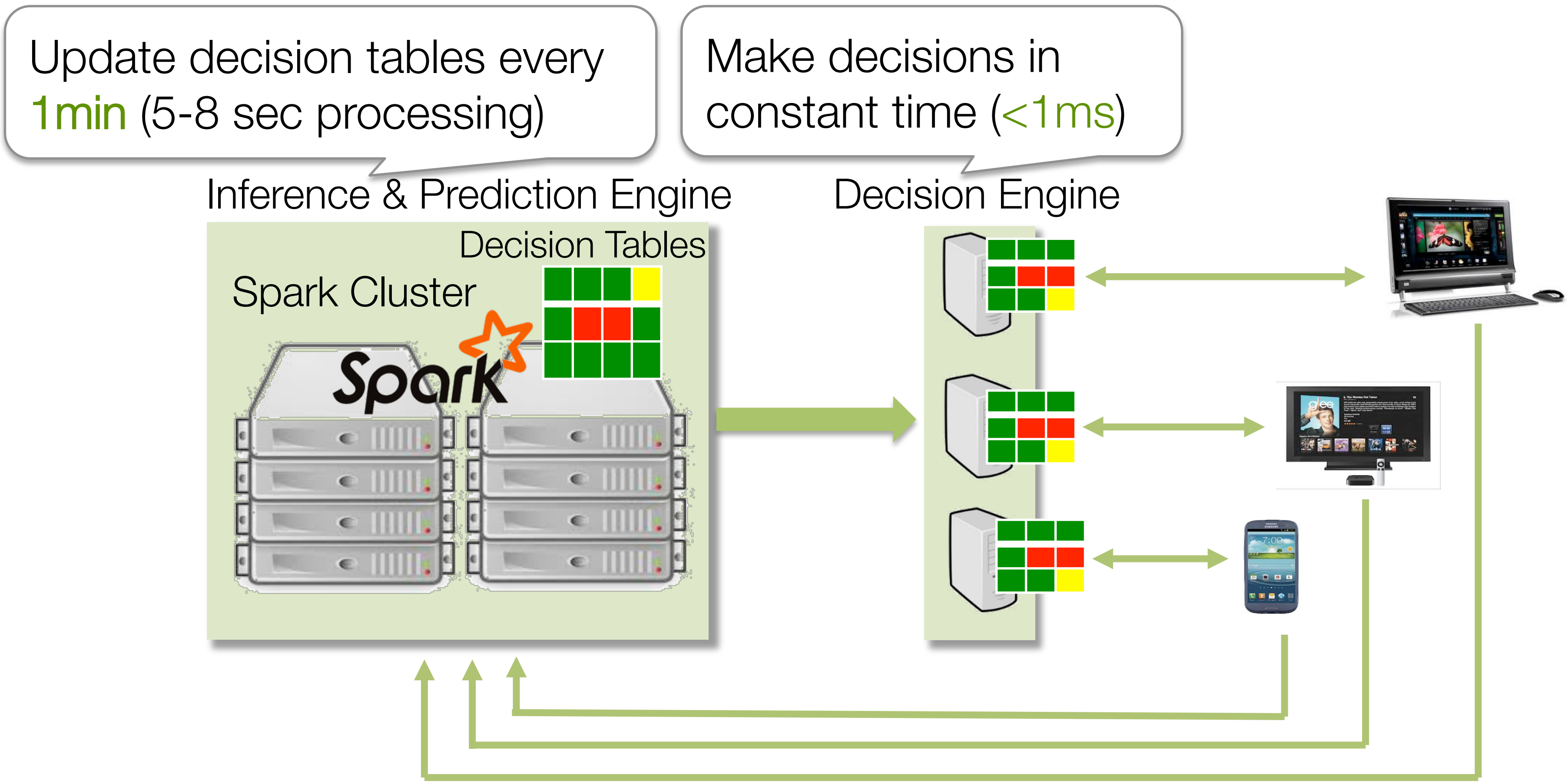
- performance of new viewers at join time
- performance of existing views if we were to switch bitrate or CDN

Create a model consisting of a set of decision tables

Reactively adapting after failure too late!

CONVIVA

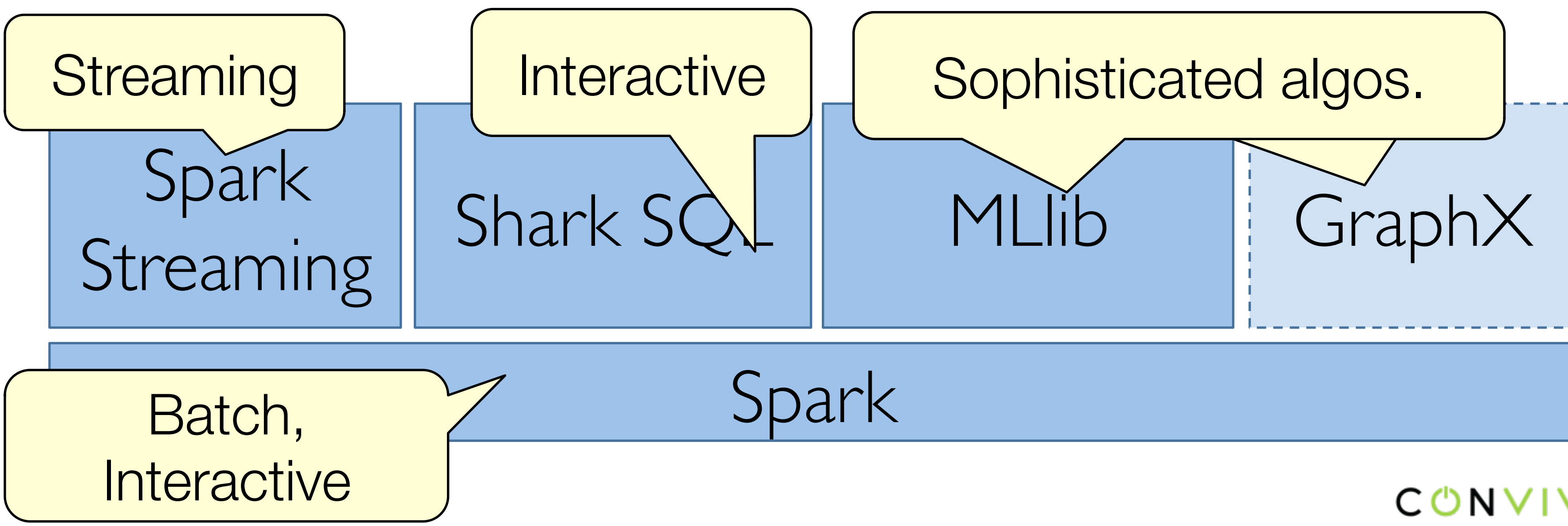
Inference and Decision Engines



CONVIVA

Spark

- In-memory computation engine
  - APIs in Scala, Python, Java
- Unifies ***batch, streaming, interactive*** computations
  - Powerful machine learning (MLlib) and soon graph (GraphX) libraries
- Used by tens of companies including Yahoo!, Intel

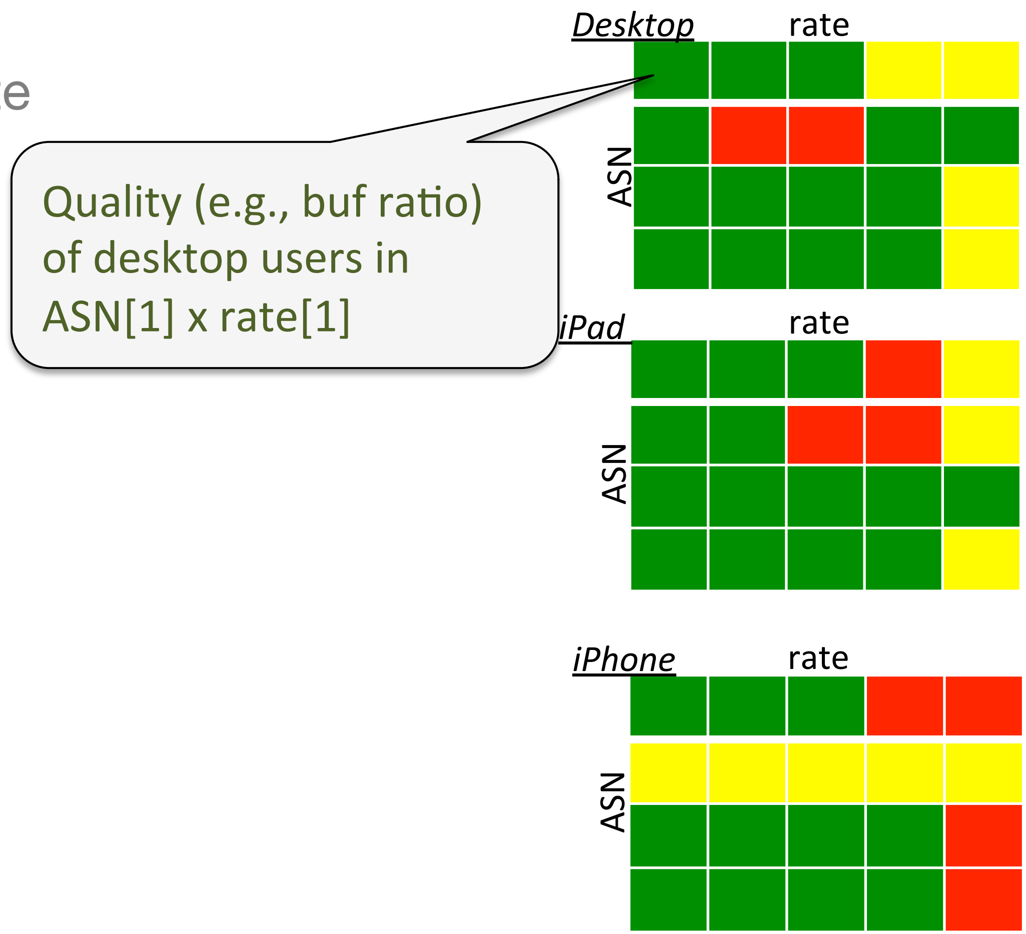


CONVIVA

Use Case 1: Best Starting Bit Rate for Single CDN

Pick the best starting bit rate based on decision tables…

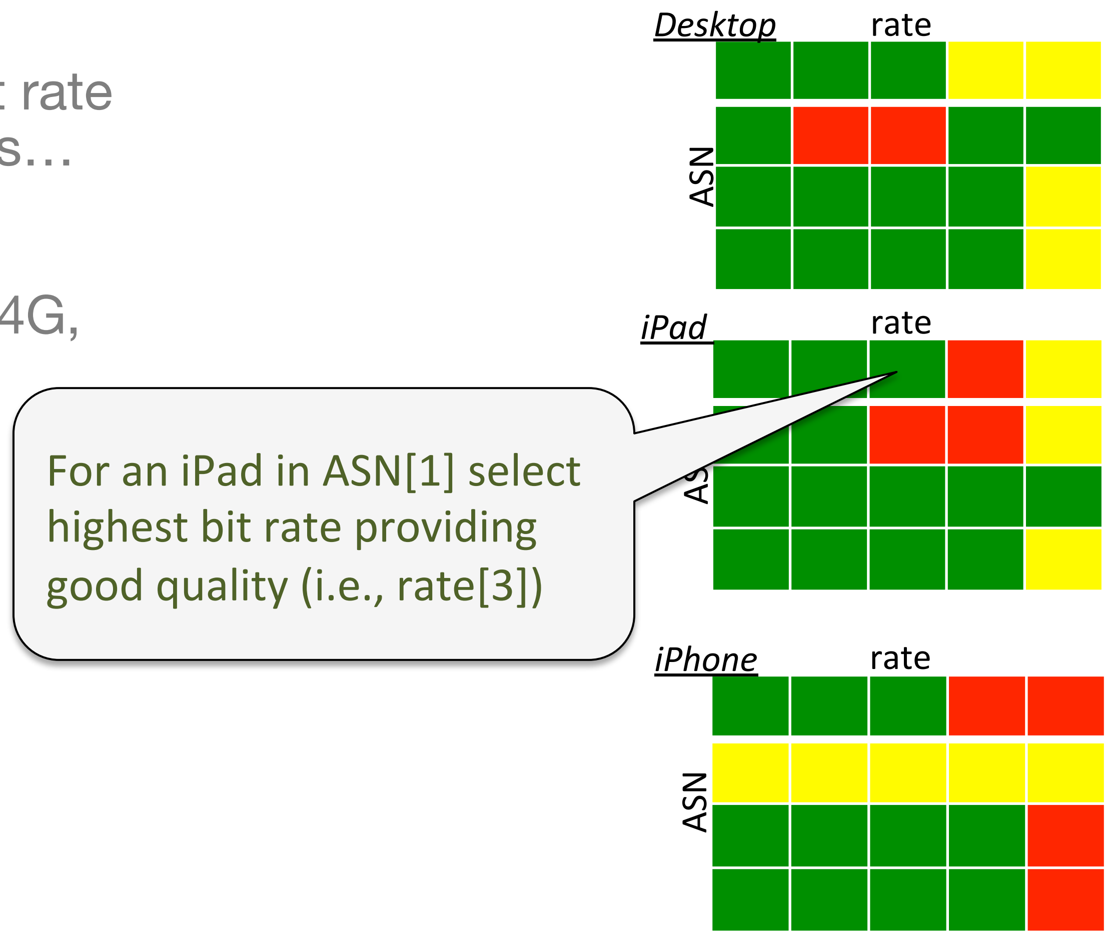
- Device
- Connection type (3G, 4G, wifi)
- Geo (DMA)
- ASN
- Protocol
- Player version
- Etc.



Use Case 1: Best Starting Bit Rate for Single CDN

Pick the best starting bit rate based on decision tables…

- Device
- Connection type (3G, 4G, wifi)
- Geo (DMA)
- ASN
- Protocol
- Player version
- Etc.

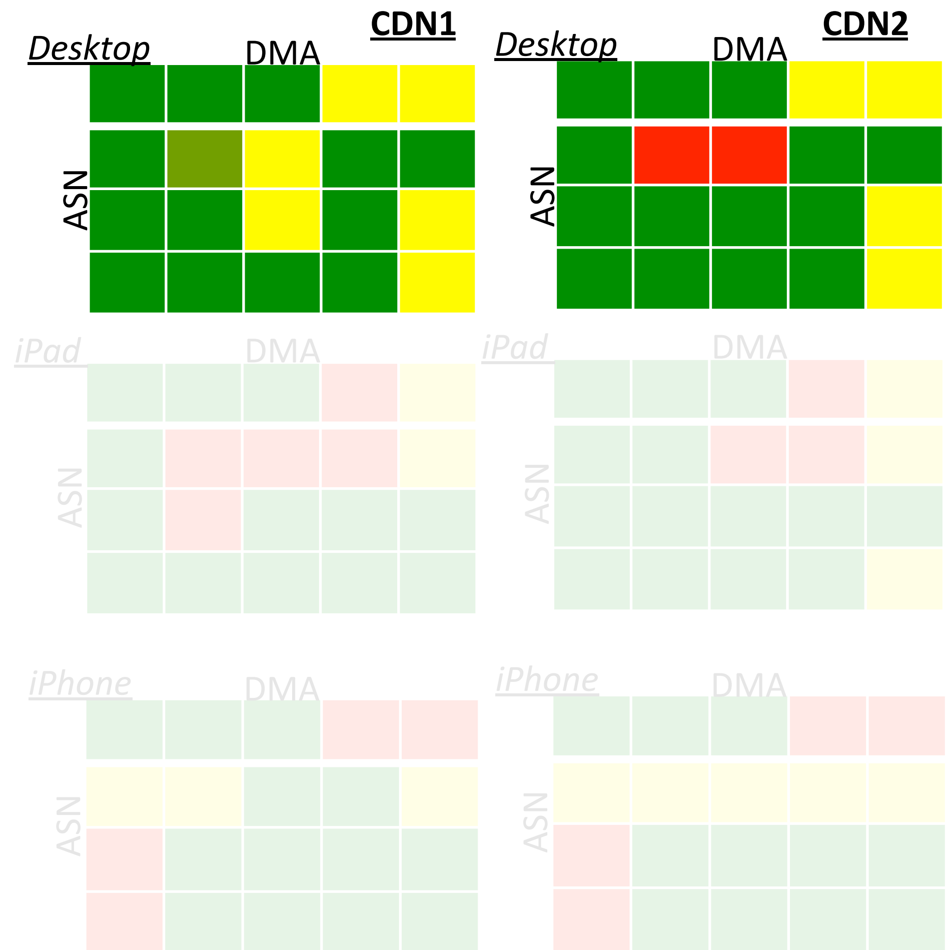


CONVIVA

Use Case 2: Best Starting CDN for Multi-CDNs

Pick the best CDN based on decision tables…

- Device
- Connection type (3G, 4G, wifi)
- Geo (DMA)
- ASN
- Protocol
- Player version
- Etc.

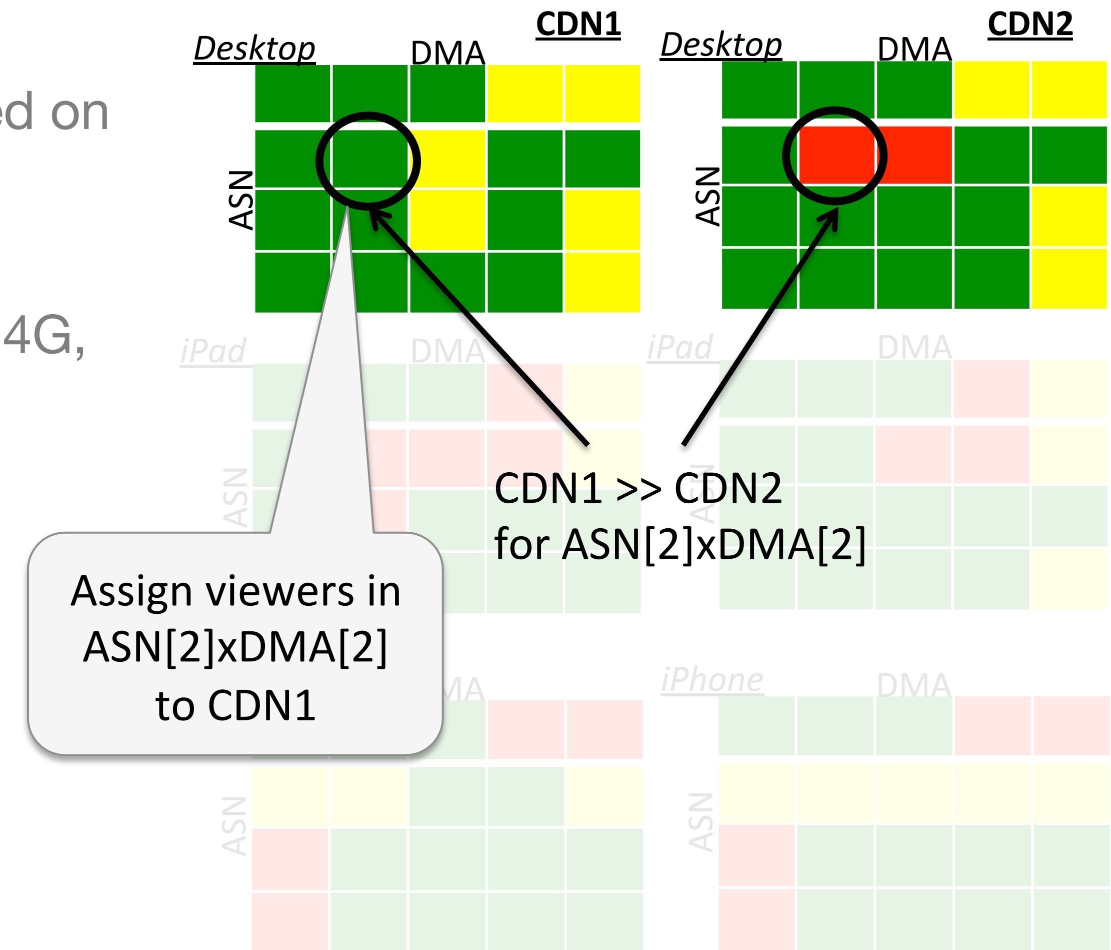


CONVIVA

Use Case 2: Best Starting CDN for Multi-CDN

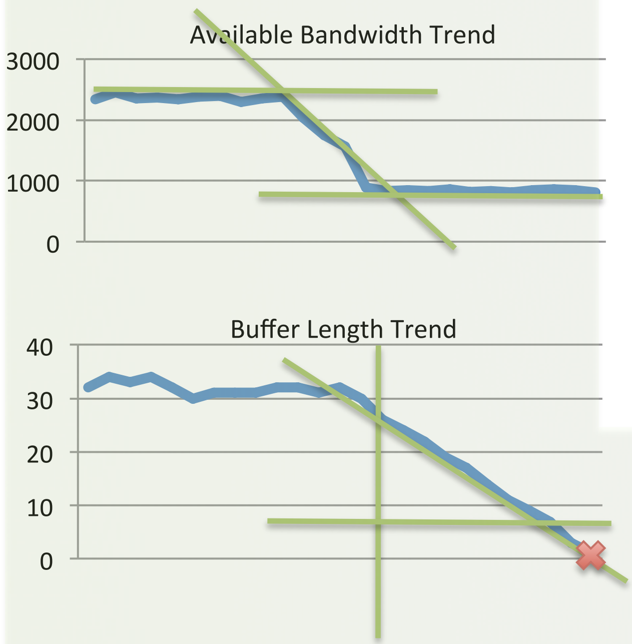
Pick the best CDN based on decision tables…

- Device
- Connection type (3G, 4G, wifi)
- Geo (DMA)
- ASN
- Protocol
- Player version
- Etc.

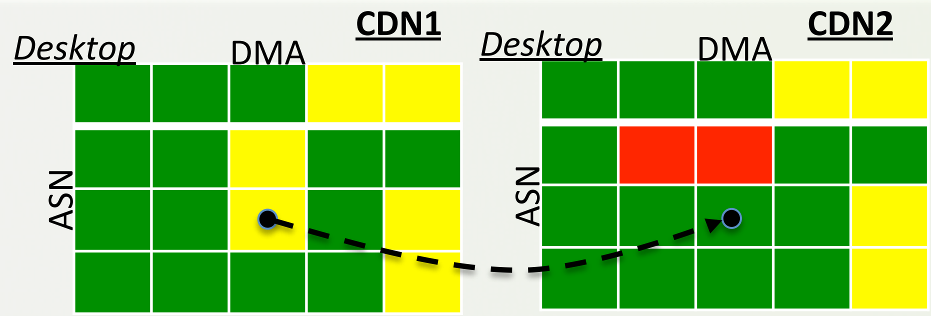


CONVIVA

Use Case 3: CDN Switch on Quality Degradation



Detect and track buffer drop and bandwidth trends and take action to prevent buffering



Client centric predictive algorithms track recent trends in

available bandwidth,
buffer length,
rendering performance …

and predict quality problems

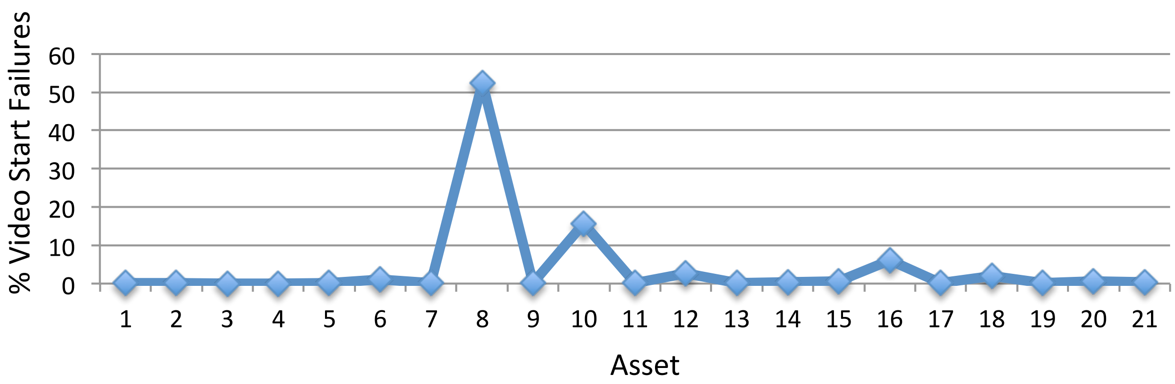
Global Inference & Prediction algorithms track global trends by

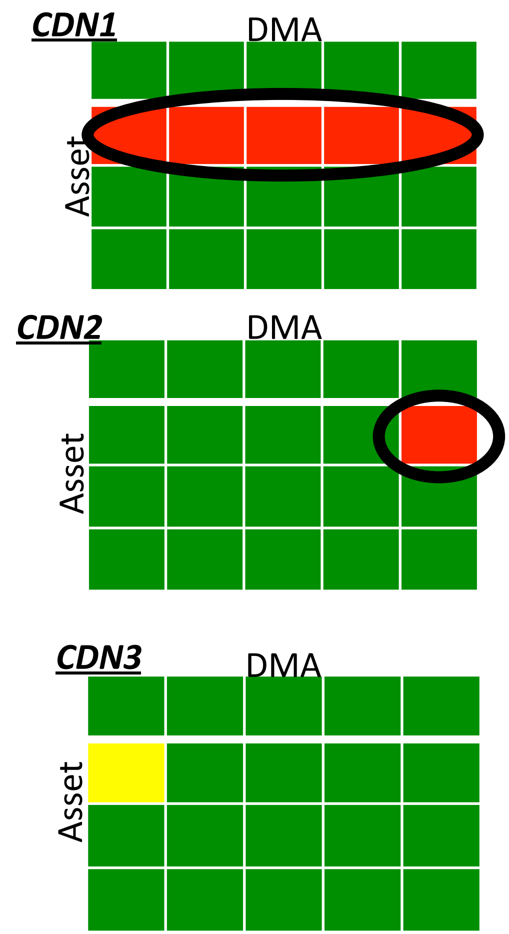
geography,
network (ASN),
CDN…

identify anomalies and predict quality of views

CONVIVA

Use Case 4: Asset Publishing and Caching Issues





Global inference algorithms track individual asset failures by

device

geography,

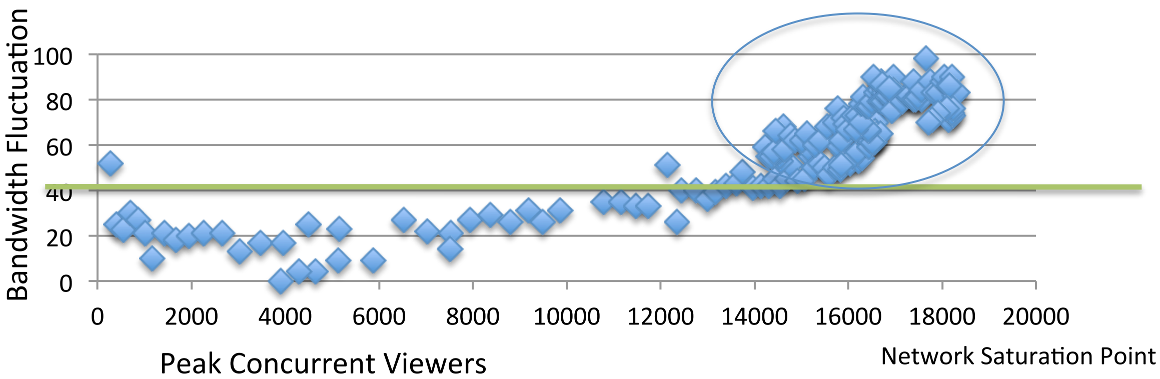
network (ASN),

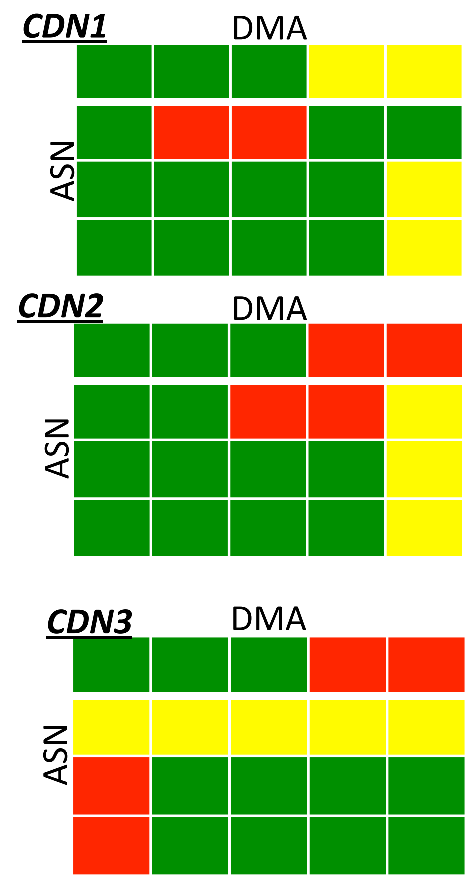
CDN…

and identify any regional anomalies

CONVIVA

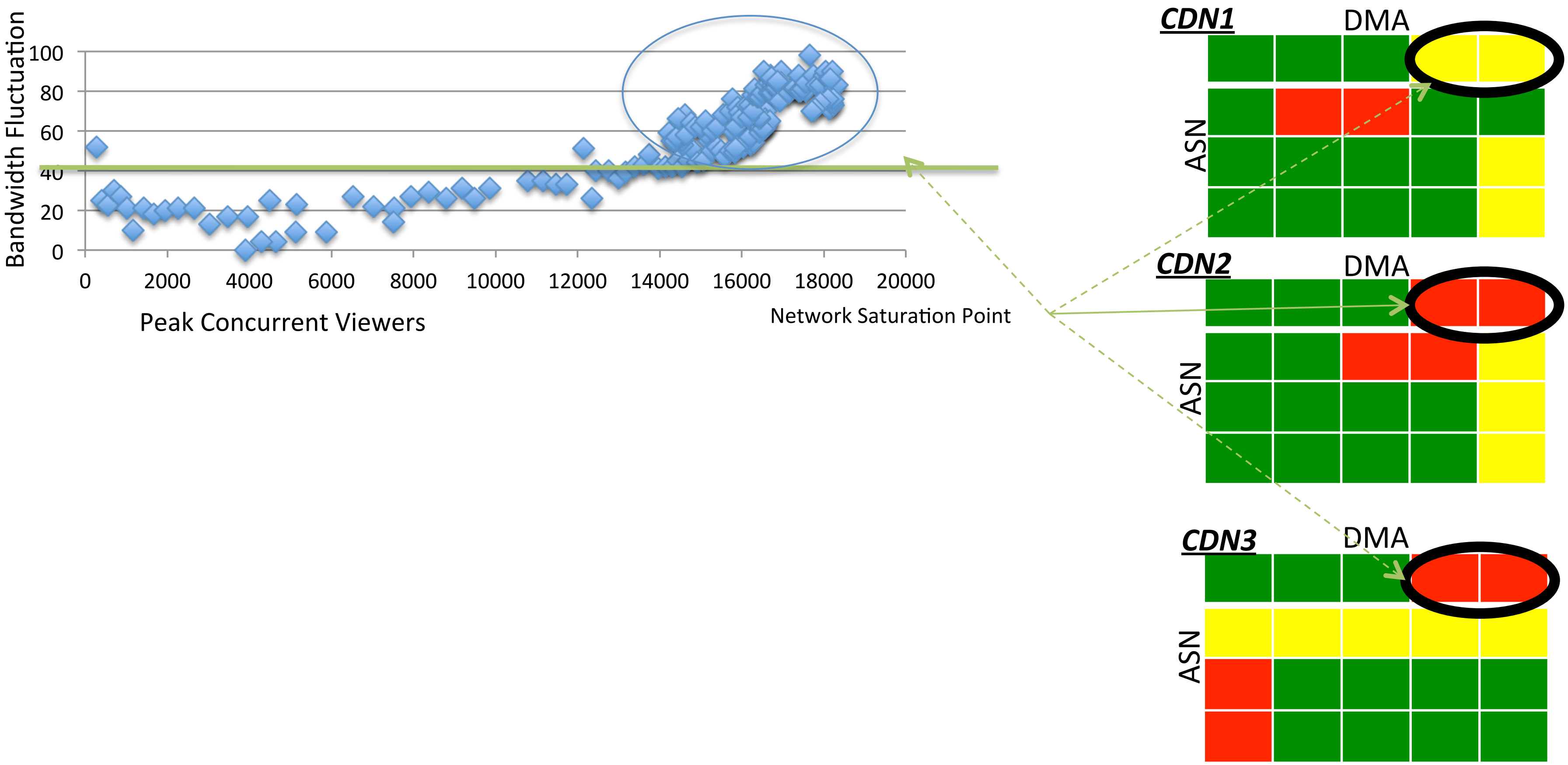
Use Case 5: ISP Saturation





This ASN/DMA is saturated on all three CDNs → Don't switch CDN. Reduce bit rates and maintain

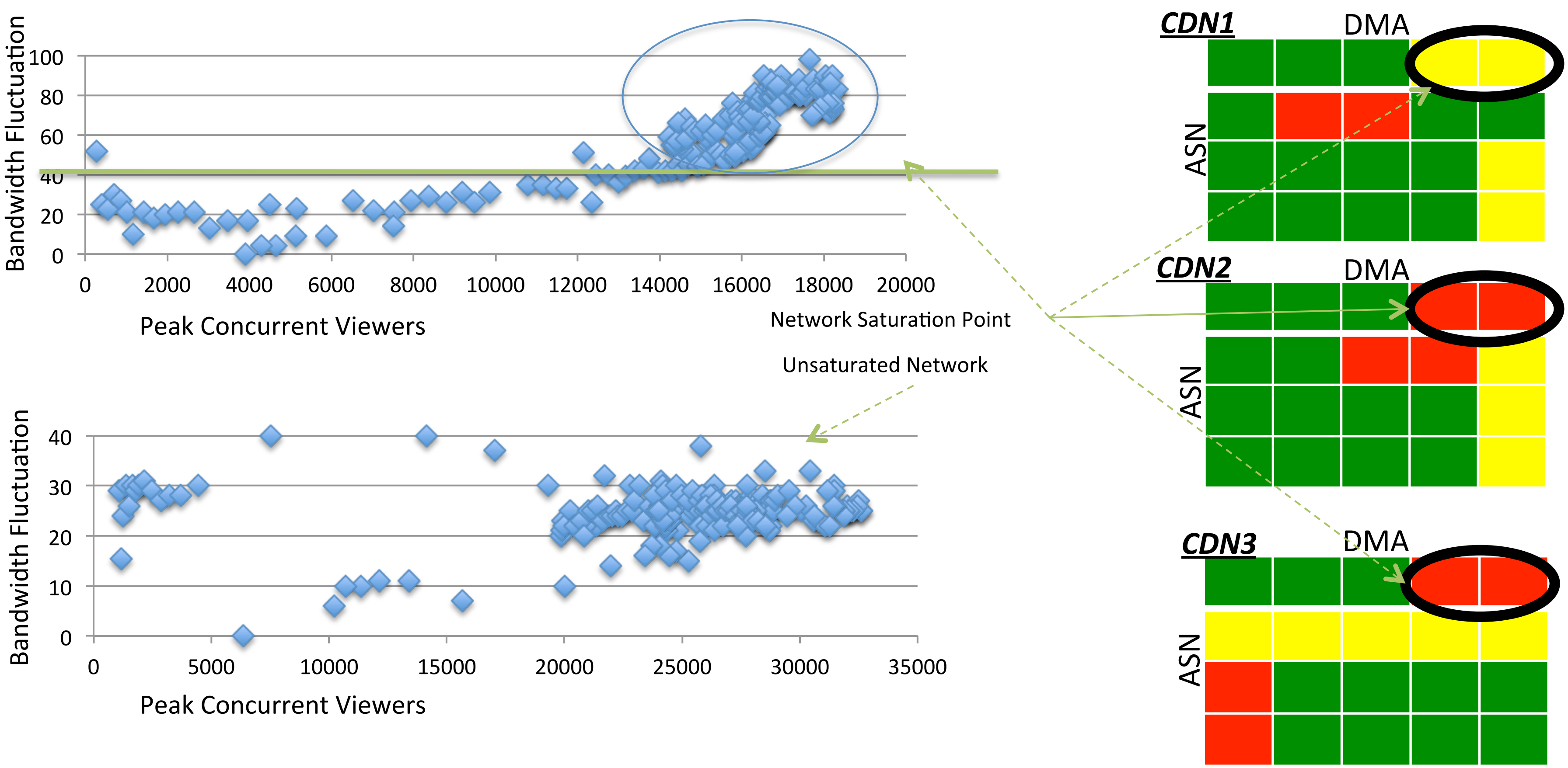
Use Case 5: ISP Saturation



This ASN/DMA is saturated on all three CDNs → Don't switch CDN. Reduce bit rates and maintain

CONVIVA

Use Case 5: ISP Saturation

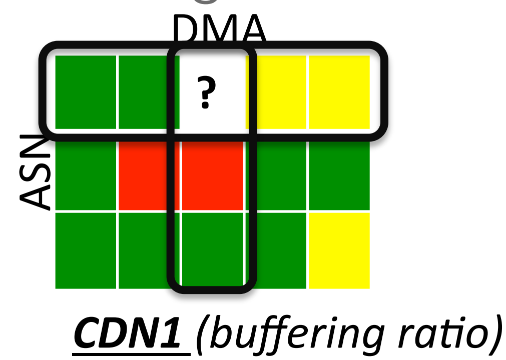


This ASN/DMA is saturated on all three CDNs → Don't switch CDN. Reduce bit rates and maintain

Challenges

- What happens if a partition doesn't has enough data?

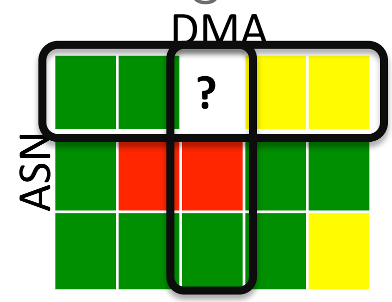
1) Spatial aggregation
   - Which partition use for prediction?



Challenges

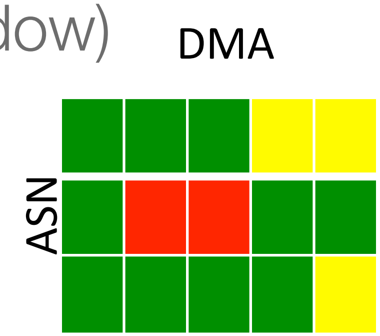
- What happens if a partition doesn't has enough data?

1) Spatial aggregation
   - Which partition use for prediction?



CDN1 *(buffering ratio)*

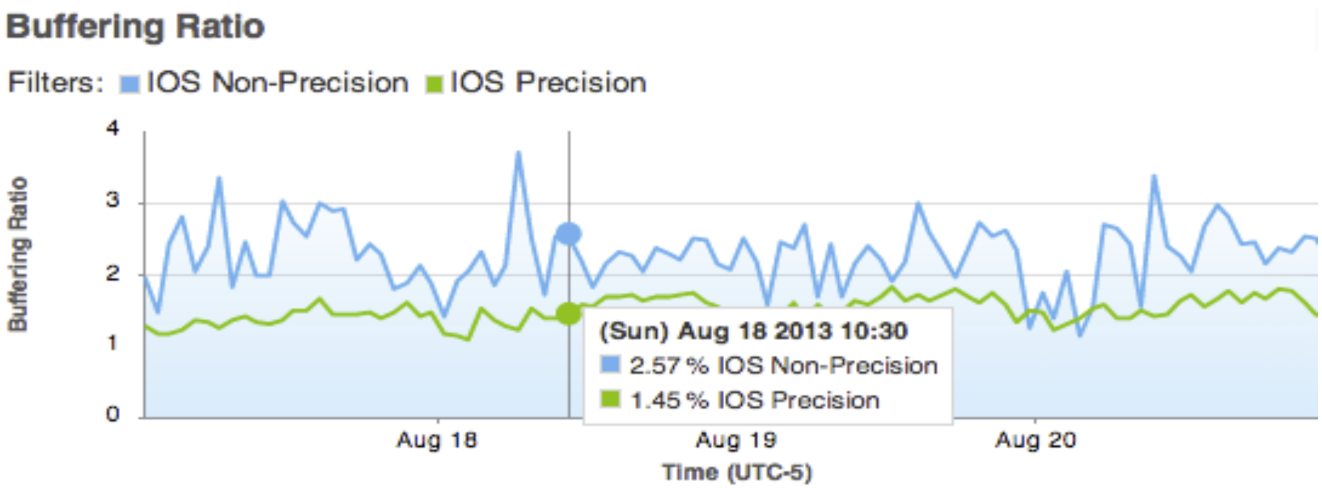
2) Temporal aggregation (increase window)
   - What window size?

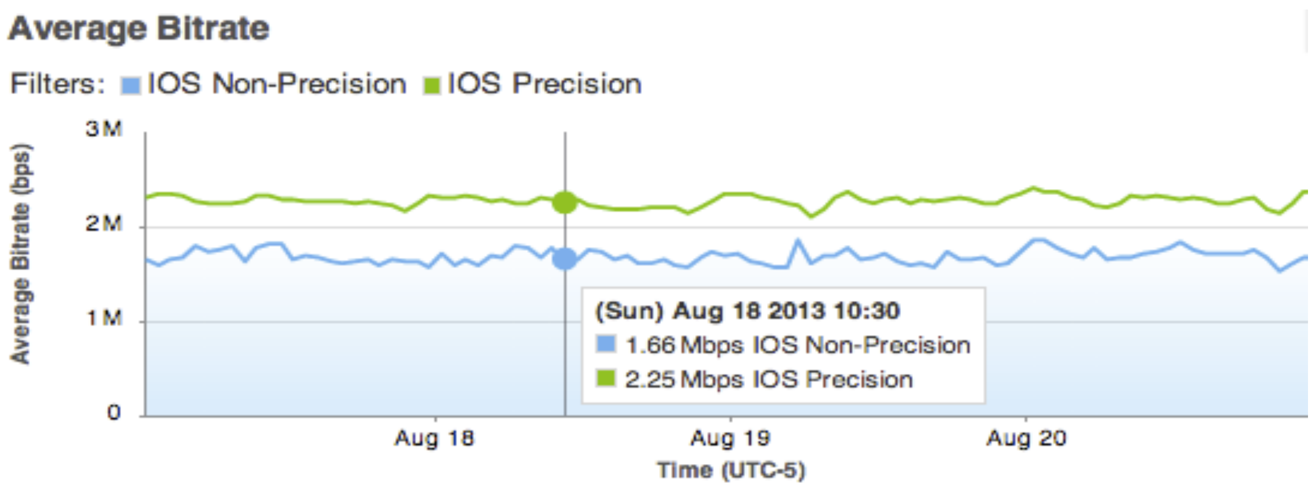


CDN1 *(buffering ratio)*

CONVIVA

Example 1: CDN and Bit Rate Selection





CONVIVA

Example 1: Impact on CDN Usage

- ASN 7922 (Comcast) – for all of Aug 19th

Metric	Buffering Ratio	Avg. Bitrate	Video failures
CDN1	1.26 %	2270 K	9.3%
CDN2	1.08 %	2668 K	8.12%
CDN3	1.47 %	2451	9.9%

Precision picks CDN2 64% of the time

- ASN 20057 (AT&T Wireless) for all of Aug 19th

Metric	Buffering Ratio	Avg. Bitrate	Video failures
CDN1	2.12 %	1832 K	10.7%
CDN2	2.46 %	1874 K	11.0%
CDN3	2.09 %	1830 K	9.6%

Precision picks CDN1 & CDN3 74% of the time

CONVIVA

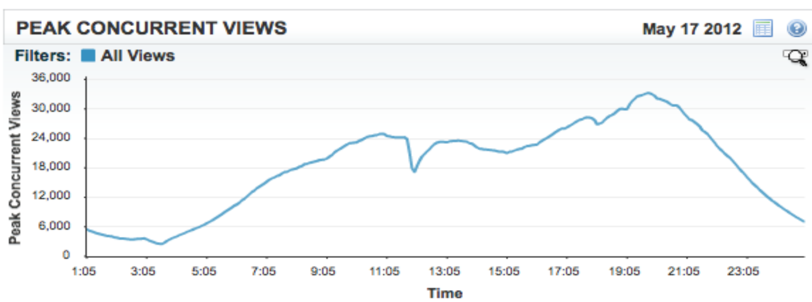
Example 1: Aggregate over 5 Days

Metric	Non-Precision	Precision	Precision Improvement
Buffering Ratio	2.3 %	1.5 %	33% ↓
Average Bitrate	1692 K	2287 K	35% ↑
Failures and Exits	11.5%	10.6 %	8% ↓
Buffering Impacted Views	13.4%	9.4 %	30% ↓

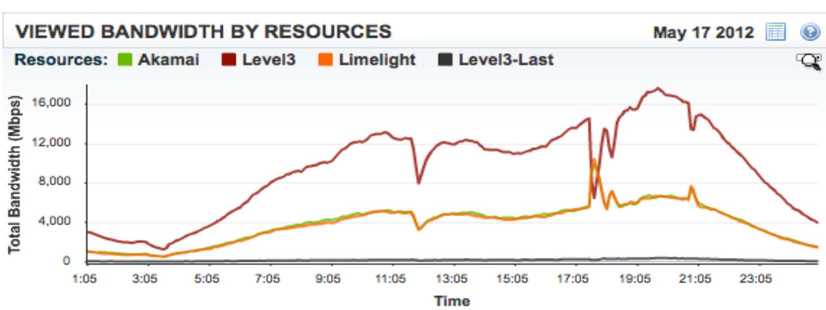
CONVIVA

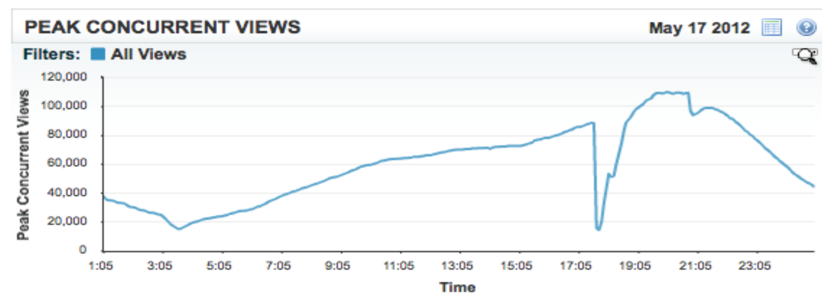
Example 2: Preserving Quality in Presence of Failures

- Precision ensures that audience quality is NOT impacted by CDN failures
- Content brands and audience are protected
- Content owners can be more aggressive in using capacity from CDN vendors

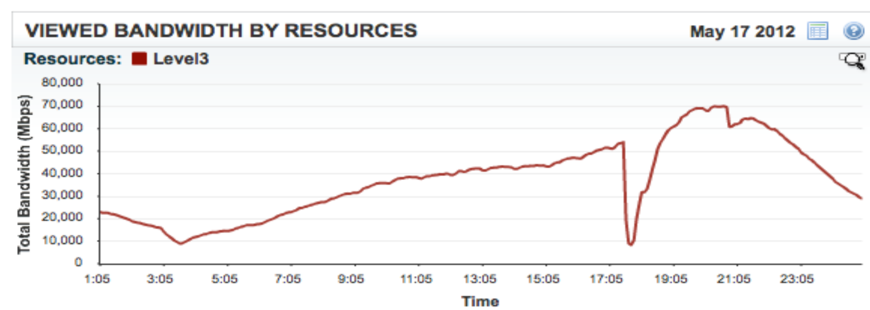


With Precision Video, CDN problem has no effect on viewers





Without Precision Video, CDN problem has big effect on viewers



Summary

- Key transition of main-stream video to the Internet

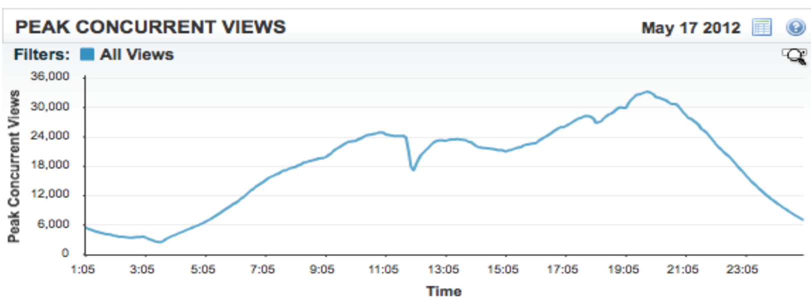
- Video quality presents opportunity and challenge
  - Premium video on big screens → zero tolerance for poor quality

- Ability to infer and predict viewer quality key to maximize quality perceived by users
  - Reacting after the fact too late!

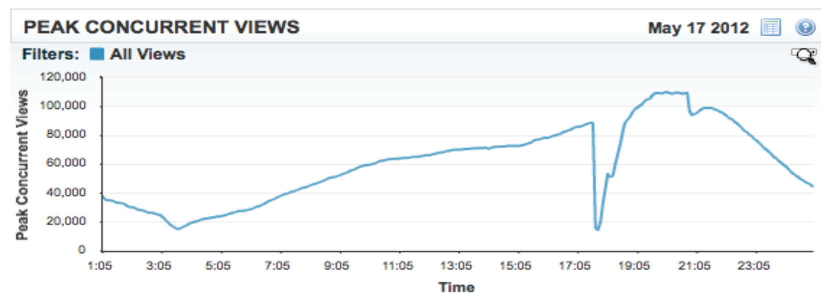
CONVIVA

Conviva Precision Protects Brands, Audience & CDN

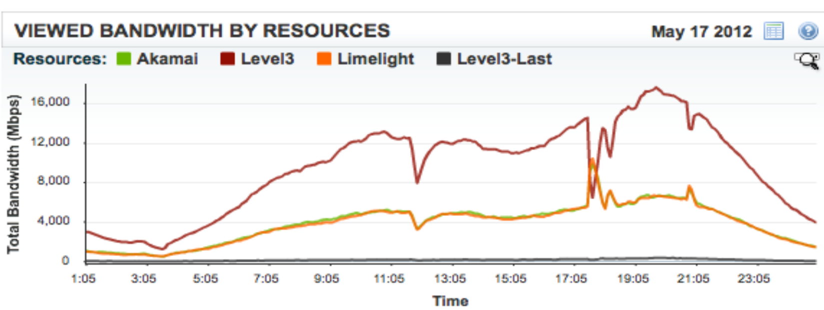
- Precision ensures that audience quality is NOT impacted by CDN failures
- Content brands and audience are protected
- Content owners can be more aggressive in using capacity from CDN vendors

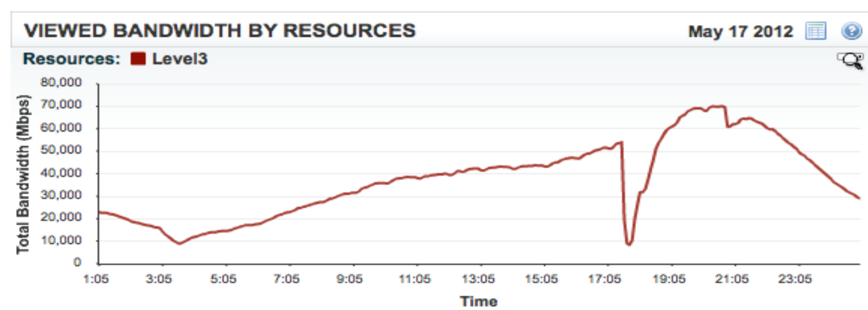


With Precision Video, CDN problem has no effect on viewers

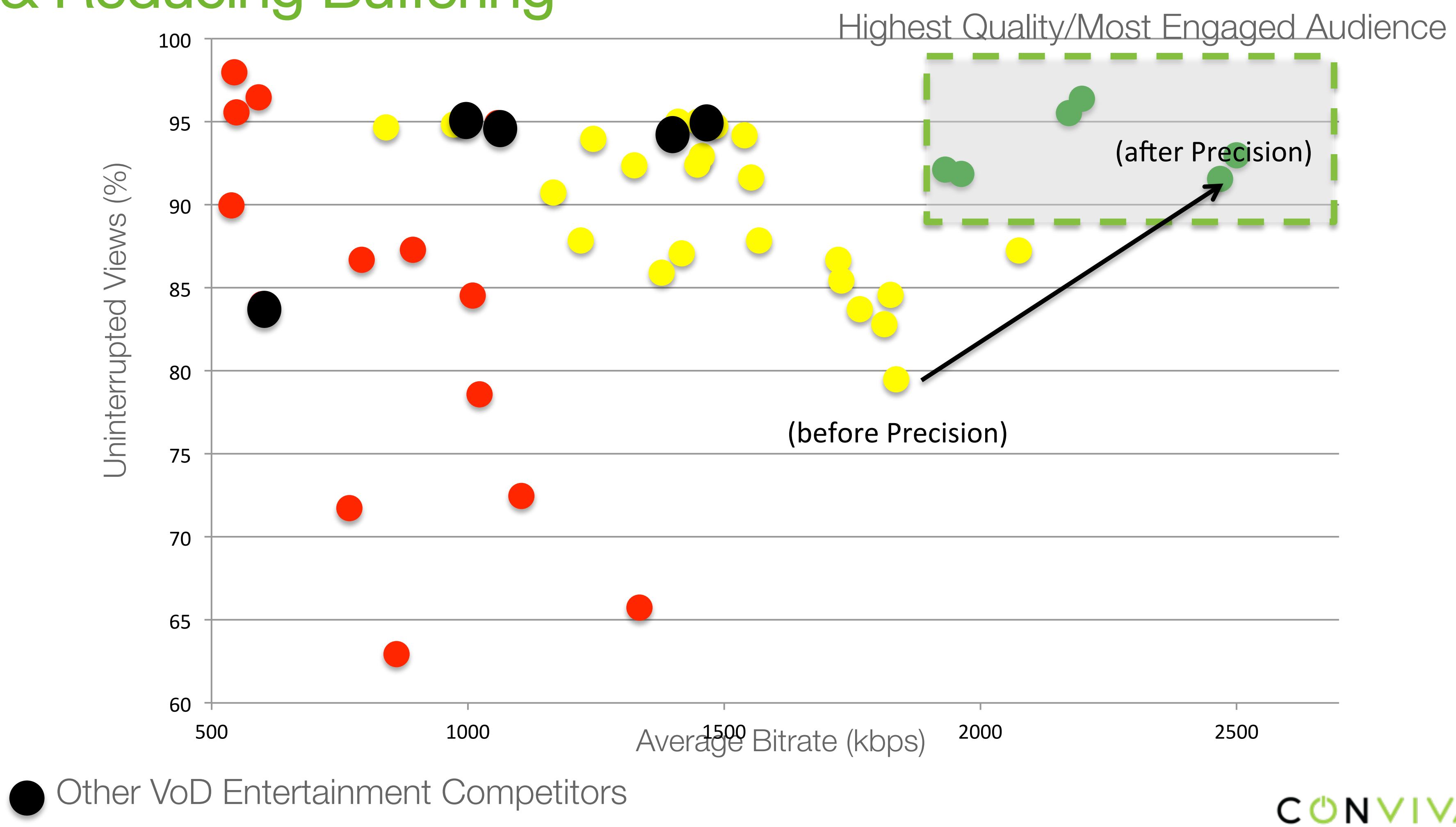


Without Precision Video, CDN problem has big effect on viewers



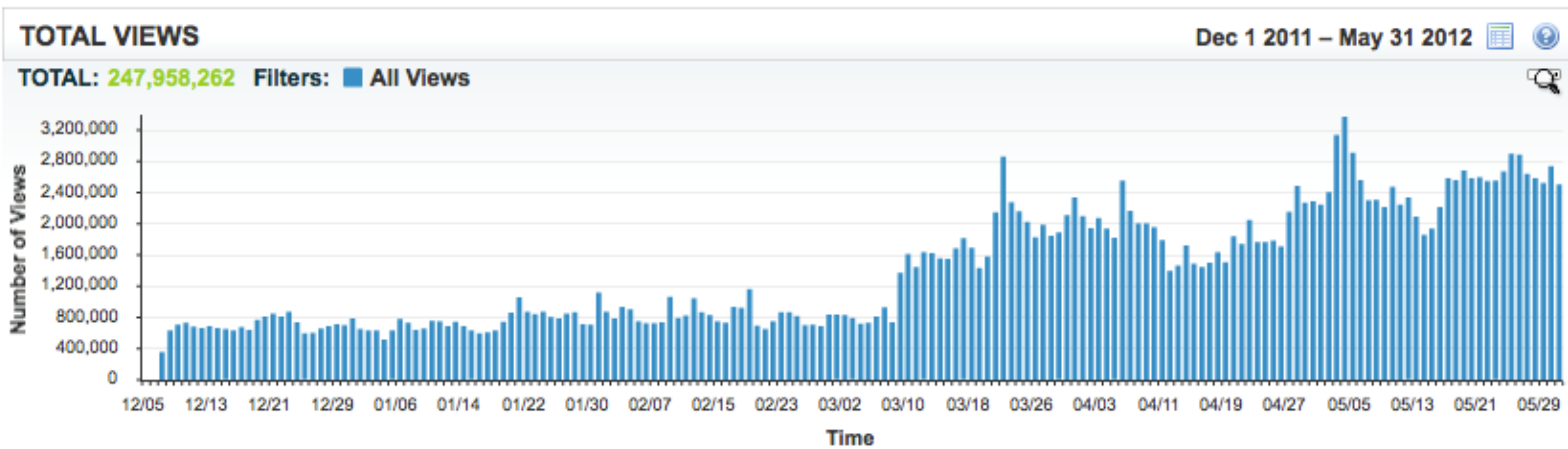


Conviva Precision: Simultaneously Increasing Resolution & Reducing Buffering

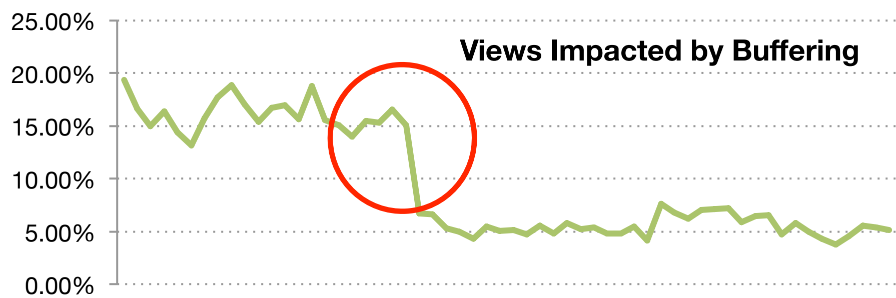


CONVIVA

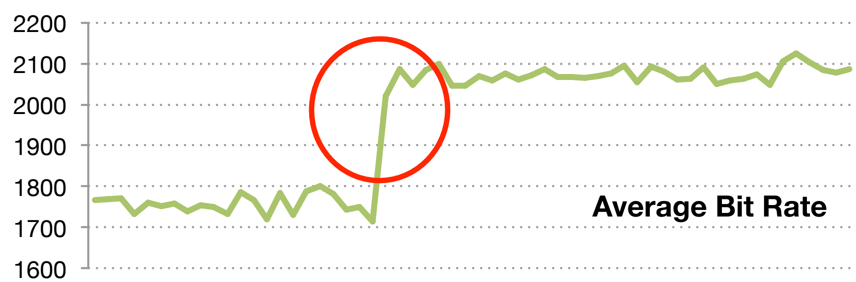
With Conviva Precision, Viewers Watch More, Come Back More Often



Lift with Precision



Reduced views impacted by buffering from 16.13% to 5.56%



Increased average bitrate from 1.7 mbps to 2.1 mbps

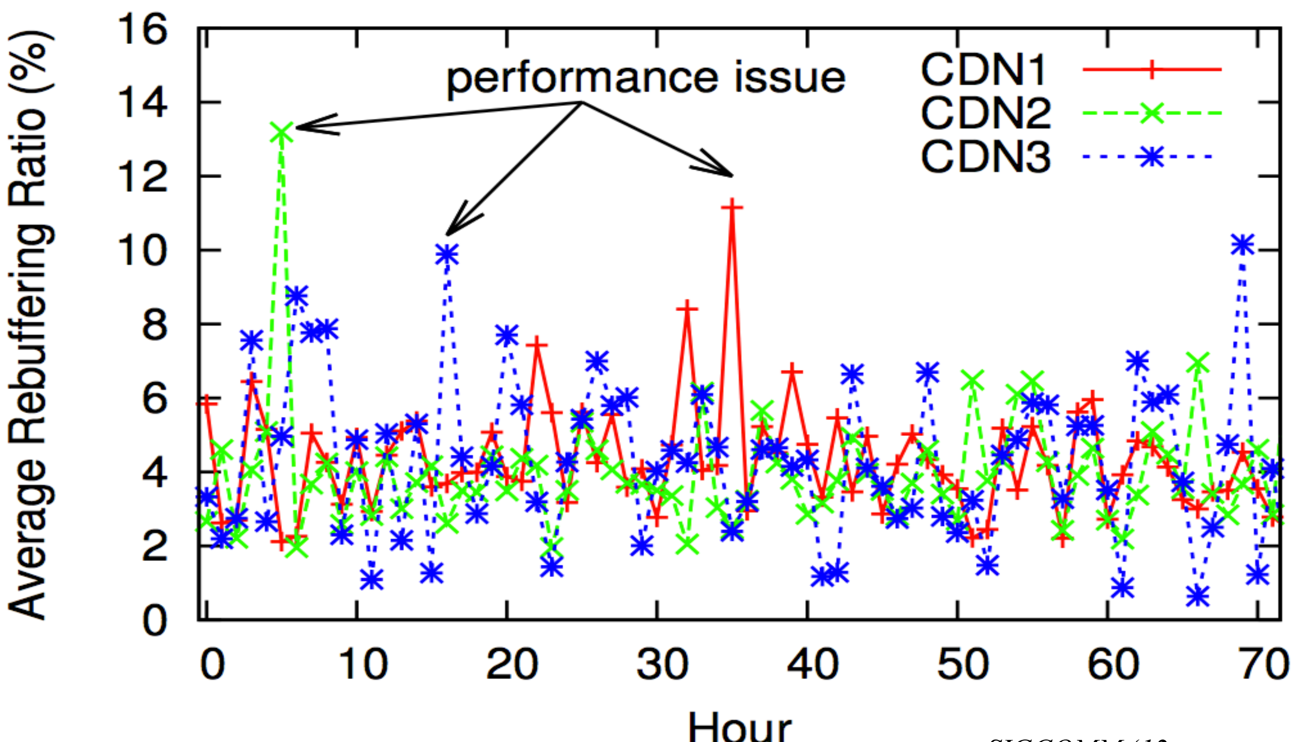
First Full Month Results

	Views	19%
	Uniques	15%
Audience	Viewed Minutes	36%
	Minutes per View	14%
	Minutes per Unique	18%

Raised engagement by 36%

The Truth

- Video delivery over the internet is hard
  - CDN variability makes it nearly impossible to deliver high quality all the time with just one CDN



SIGCOMM '12

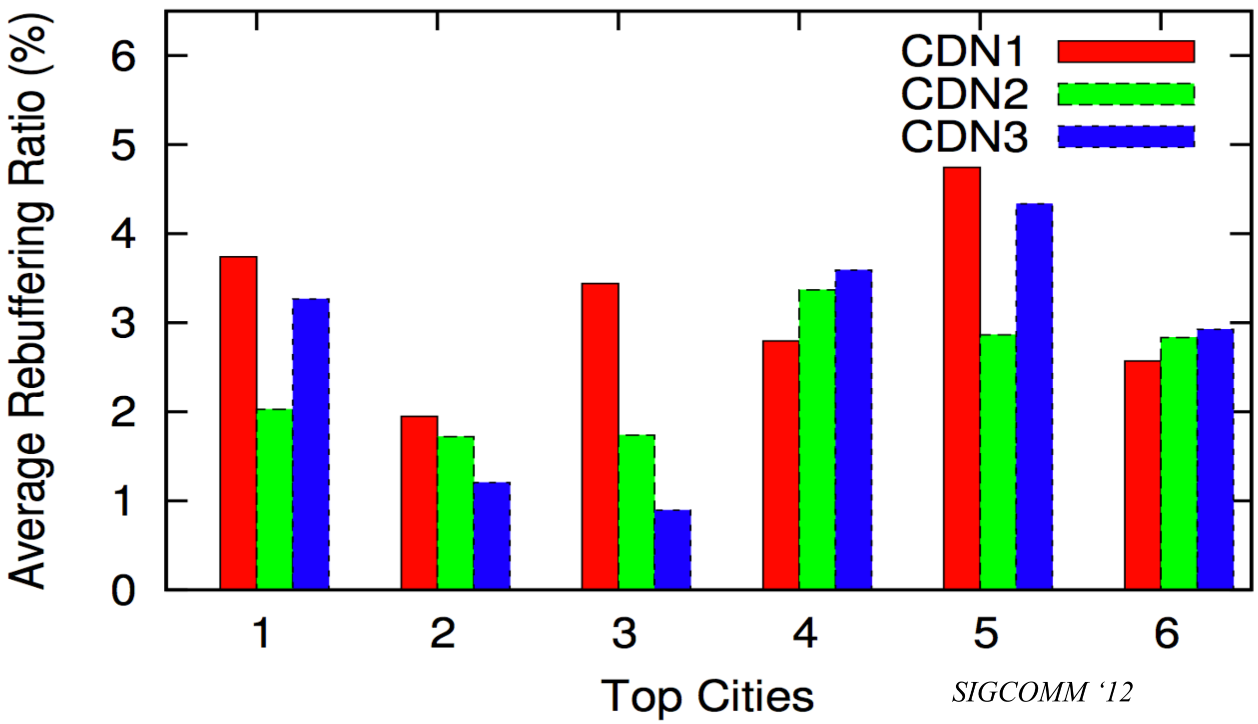
CONVIVA

The Idea

- Where there is heterogeneity, there is room for optimization
- For each viewer we want to decide what CDN to stream from
- But it's difficult to model the internet, and things can rapidly change over time
- So we will make this decision based on the real-time data that we collect

The Truth

- Video delivery over the internet is hard
  - CDN variability makes it nearly impossible to deliver high quality everywhere with just one CDN

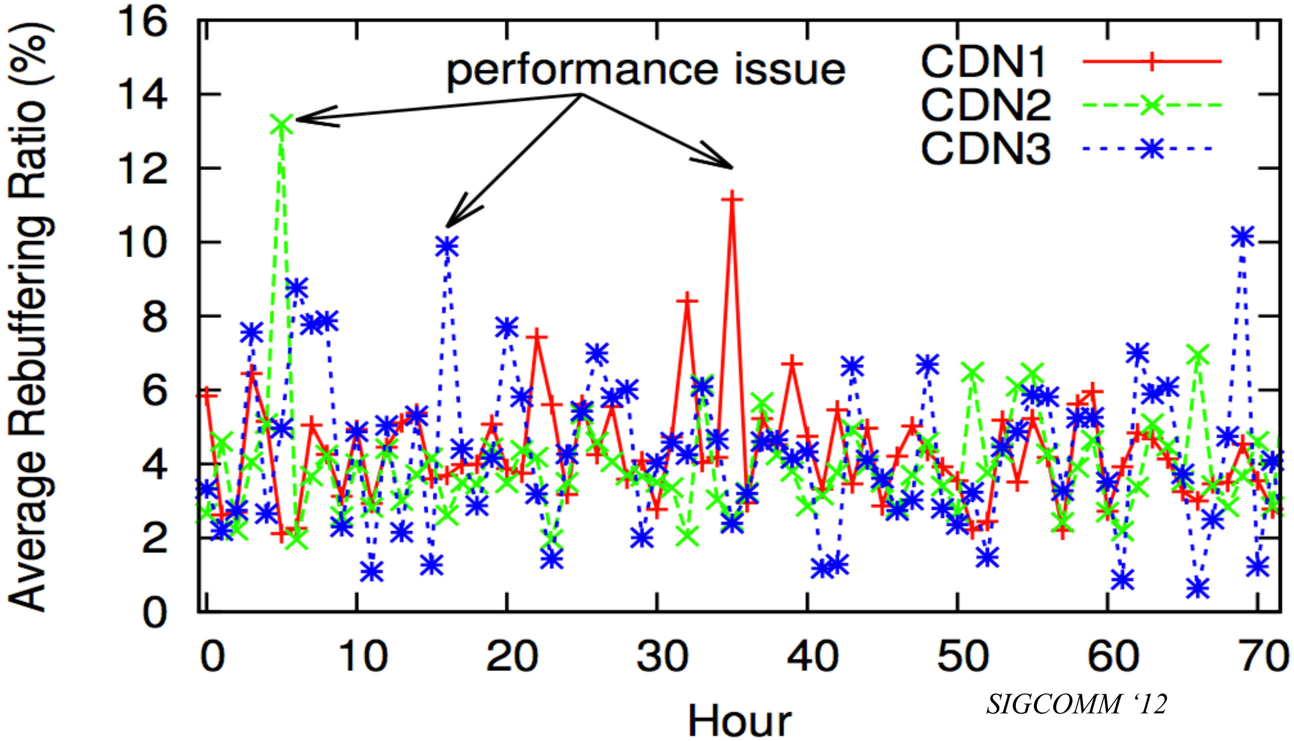


SIGCOMM '12

CONVIVA

The Truth

- Video delivery over the internet is hard
  - CDN variability makes it nearly impossible to deliver high quality all the time with just one CDN



SIGCOMM '12

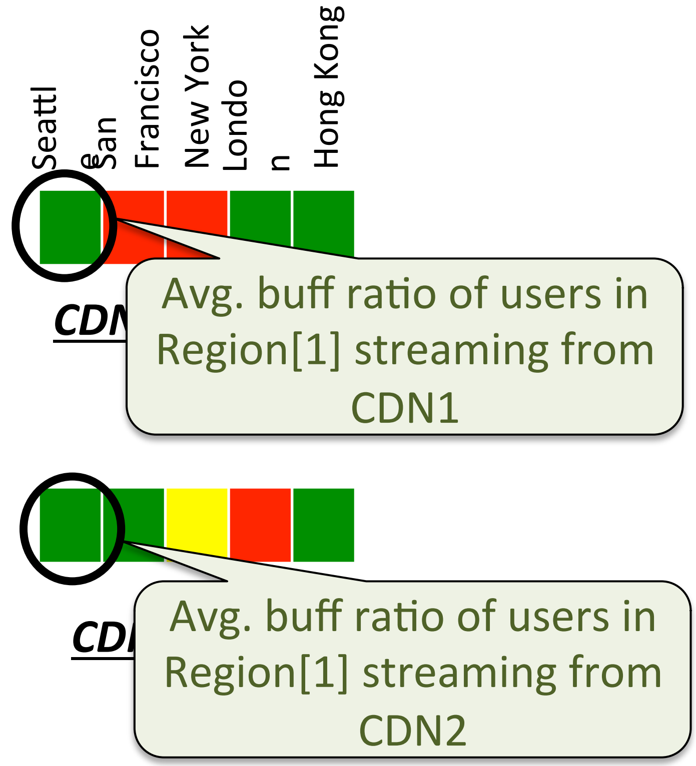
CONVIVA

The Idea

- Where there is heterogeneity, there is room for optimization
- For each viewer we want to decide what CDN to stream from
- But it's difficult to model the internet, and things can rapidly change over time
- So we will make this decision based on the real-time data that we collect

The Idea

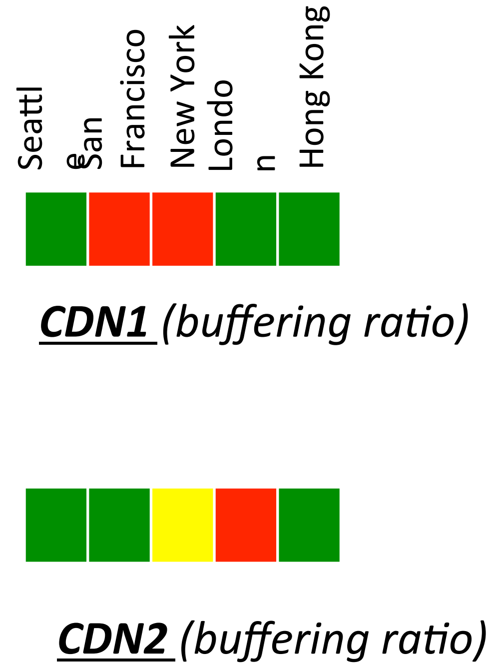
- For each CDN, partition clients by City
- For each partition compute Buffering Ratio



CONVIVA

The Idea

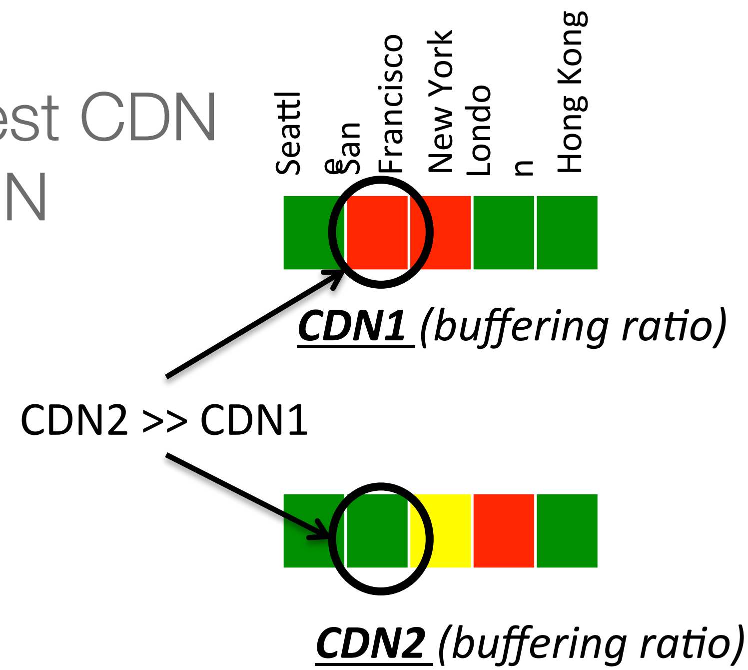
- For each partition select best CDN and send clients to this CDN



CONVIVA

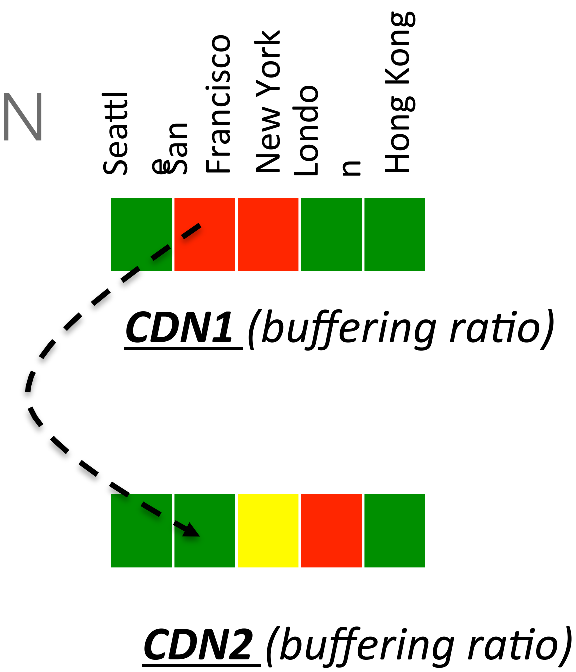
The Idea

- For each partition select best CDN and send clients to this CDN



The Idea

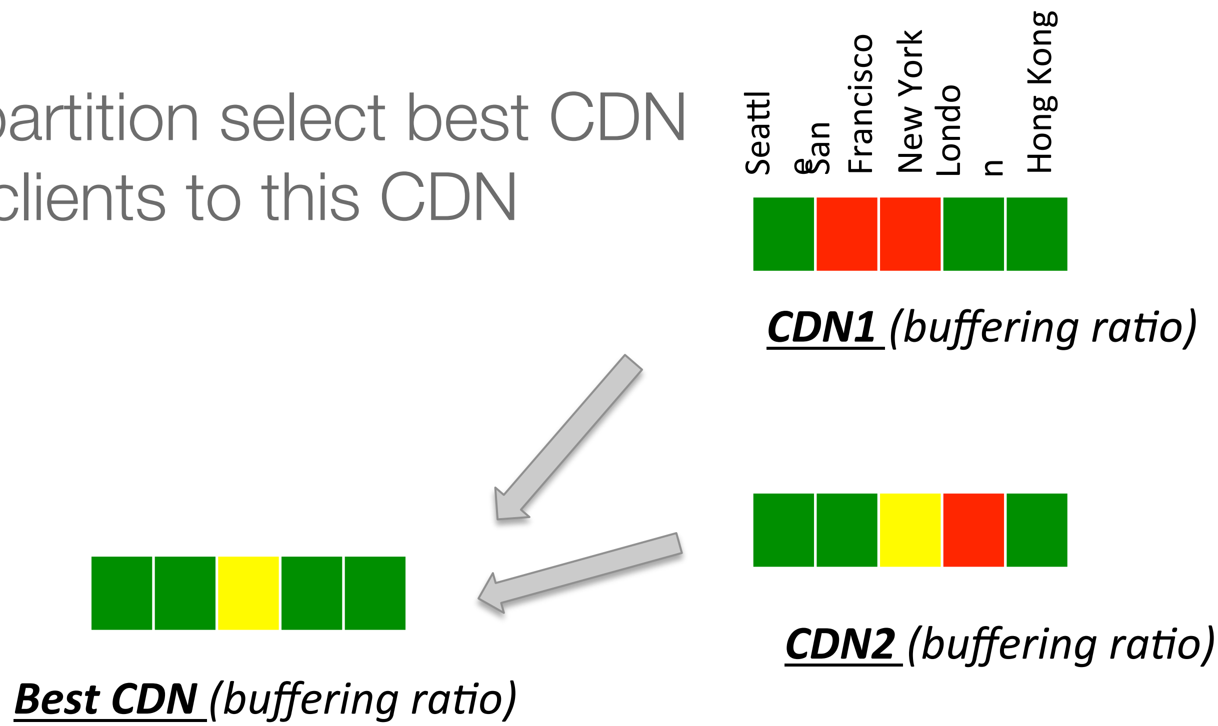
- For each partition select best CDN and send clients to this CDN



CONVIVA

The Idea

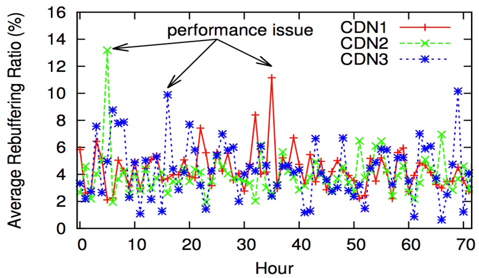
- For each partition select best CDN and send clients to this CDN



CONVIVA

The Idea

- What if there are changes in performance?



Seattle
San Francisco
New York
London
Hong Kong

***CDN1** (buffering ratio)*

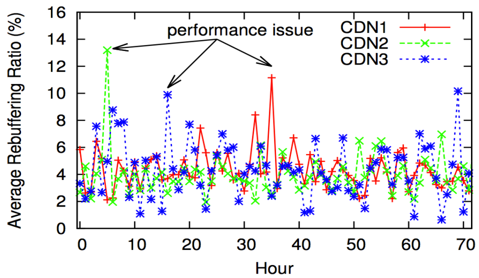
***CDN2** (buffering ratio)*

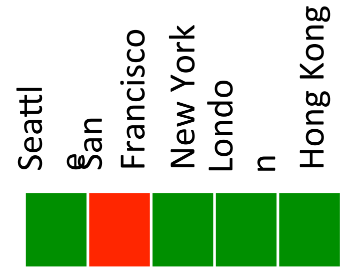
***Best CDN** (buffering ratio)*

CONVIVA

The Idea

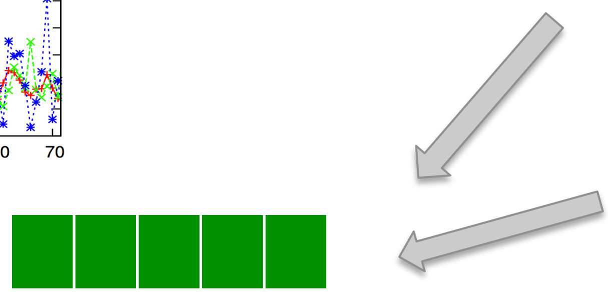
- Use online algorithm respond to changes in the network.





CDN1 *(buffering ratio)*

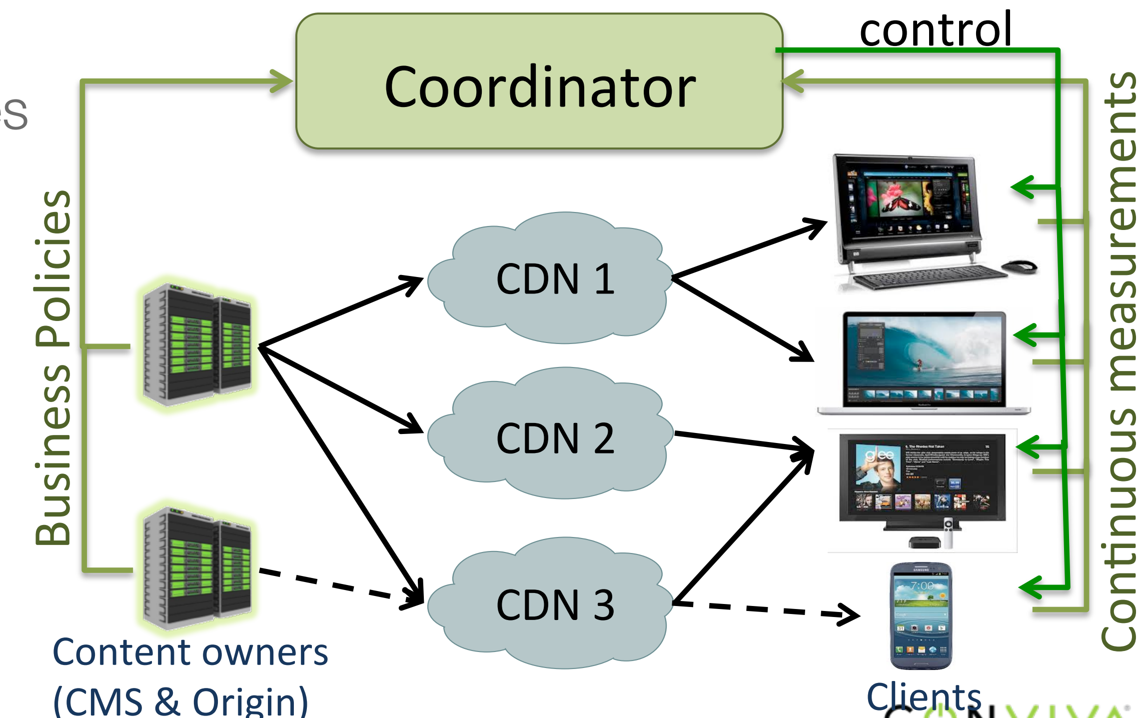
CDN2 *(buffering ratio)*



Best CDN *(buffering ratio)*

How?

- Coordinator implementing an optimization algorithm that dynamically selects a CDN for each client based on
  - Individual client
  - Aggregate statistics
  - Content owner policies
- All based on real-time data

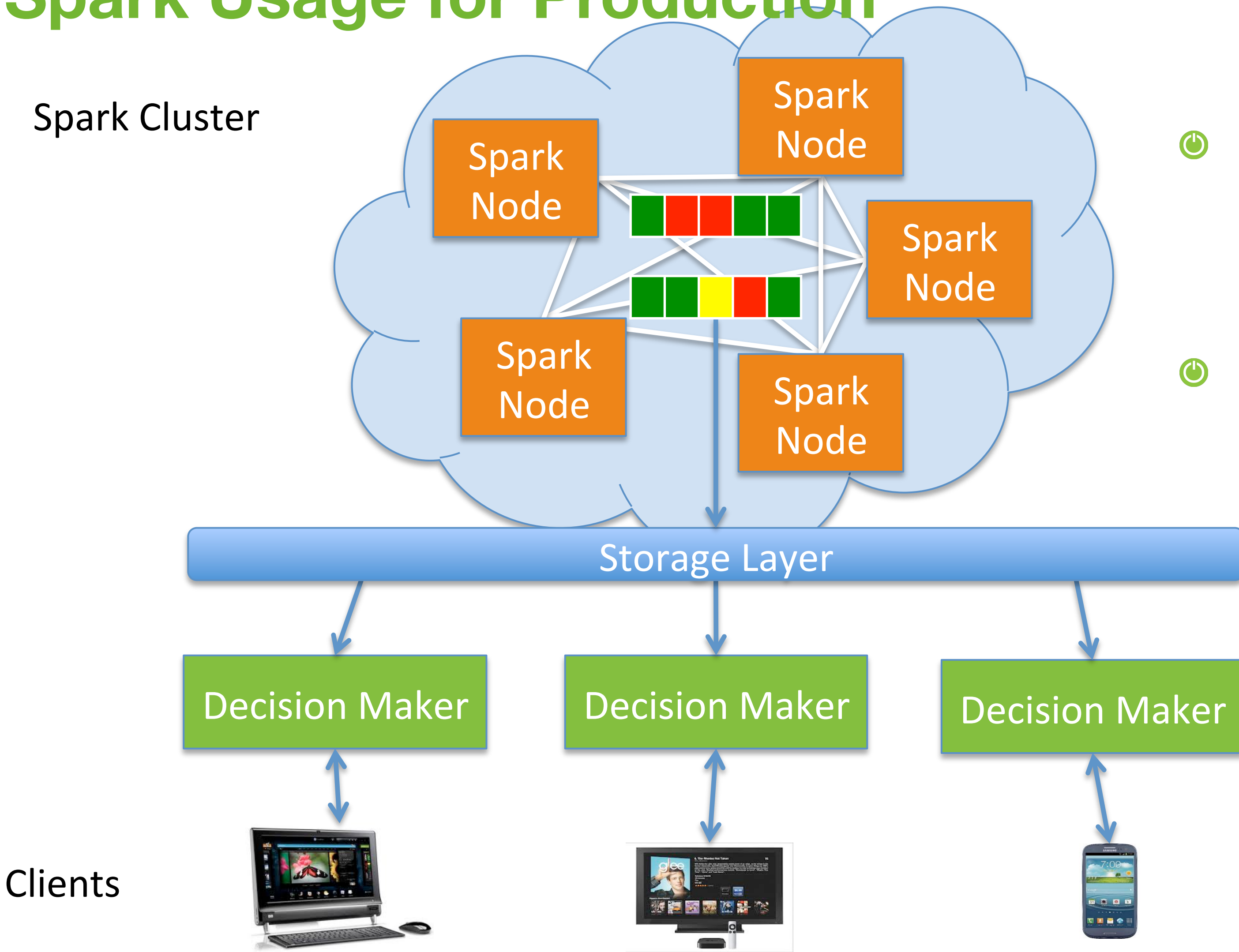


CONVIVA

What processing framework do we use?

- Twitter Storm
  - Fault tolerance model affects data accuracy
  - Non-deterministic streaming model
- Roll our own
  - Too complex
  - No need to reinvent the wheel
- Spark
  - Easily integrates with existing Hadoop architecture
  - Flexible, simple data model
  - Writing map() is generally easier than writing update()

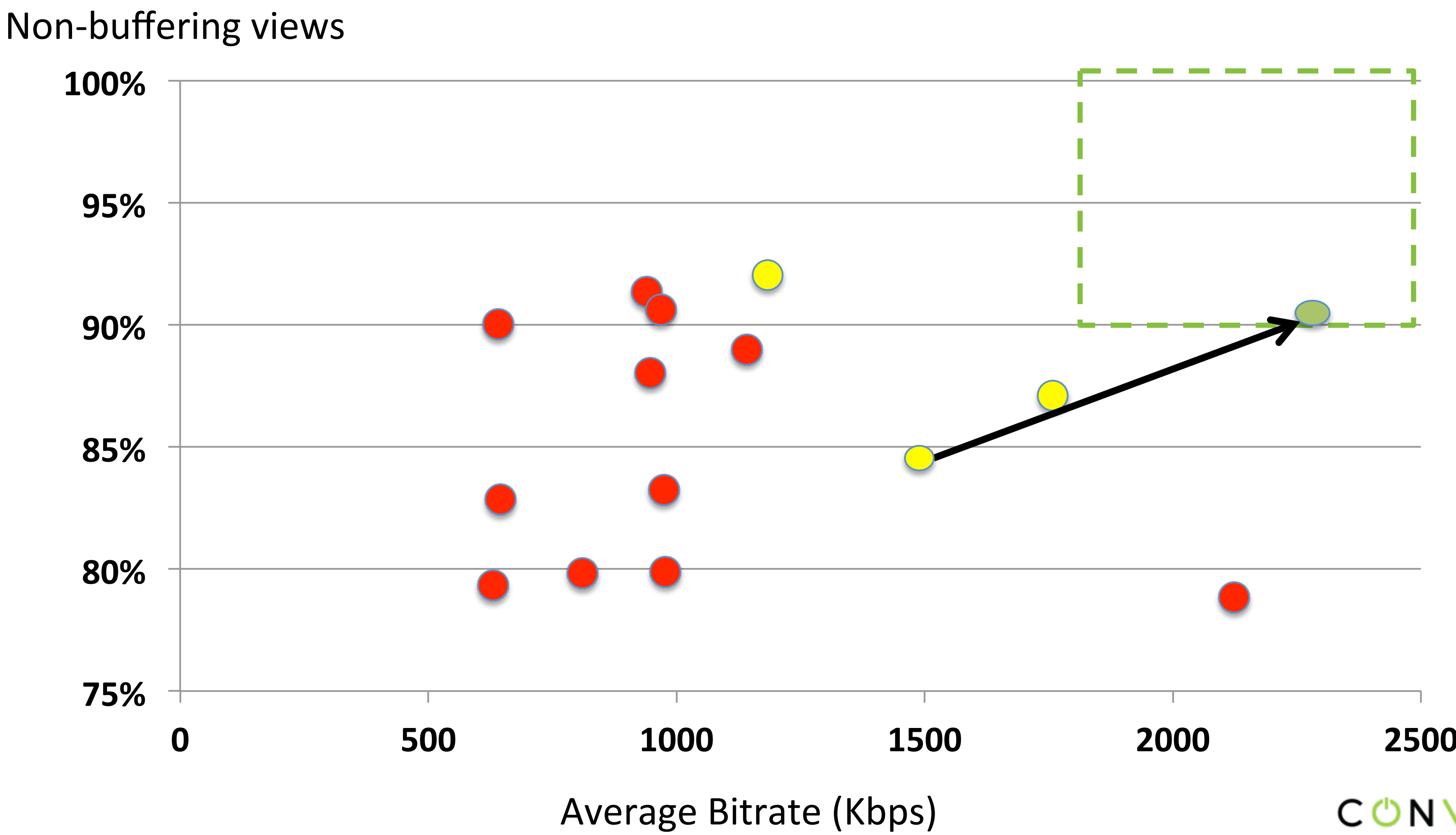
Spark Usage for Production



- Compute performance metrics in spark cluster
- Relay performance information to decision makers

CONVIVA

Results



CONVIVA

Spark's Role

- Spark development was incredibly rapid, aided both by its excellent programming interface and highly active community
- Expressive:
  - Develop complex on-line ML decision based algorithm in ~1000 lines of code
  - Easy to prototype various algorithms
- It has made scalability a far more manageable problem
- After initial teething problems, we have been running Spark in a production environment reliably for several months.

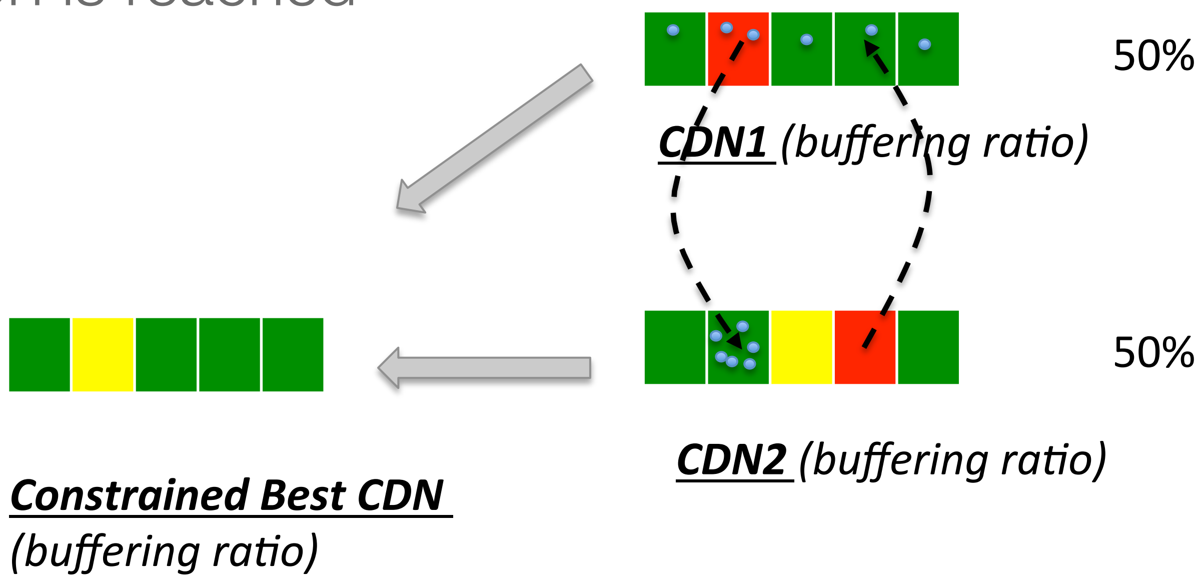
CONVIVA

Problems we faced

- Silent crashes…
- Often difficult to debug, requiring some tribal knowledge
- Difficult configuration parameters, with sometimes inexplicable results
- Fundamental understanding of underlying data model was essential to writing effective, stable spark programs

Enforcing constraints on optimization

- Imagine swapping clients until an optimal solution is reached



Enforcing constraints on top of optimization

- Solution is found *after* clients have already joined.
- Therefore we need to parameterize solution to clients already seen for online use.
- Need to compute an LP on real time data
- Spark Supported it
  - 20 LPs
  - Each with 4000 decisions variables and 350 constraints
  - 5 seconds.

CONVIVA

Tuning

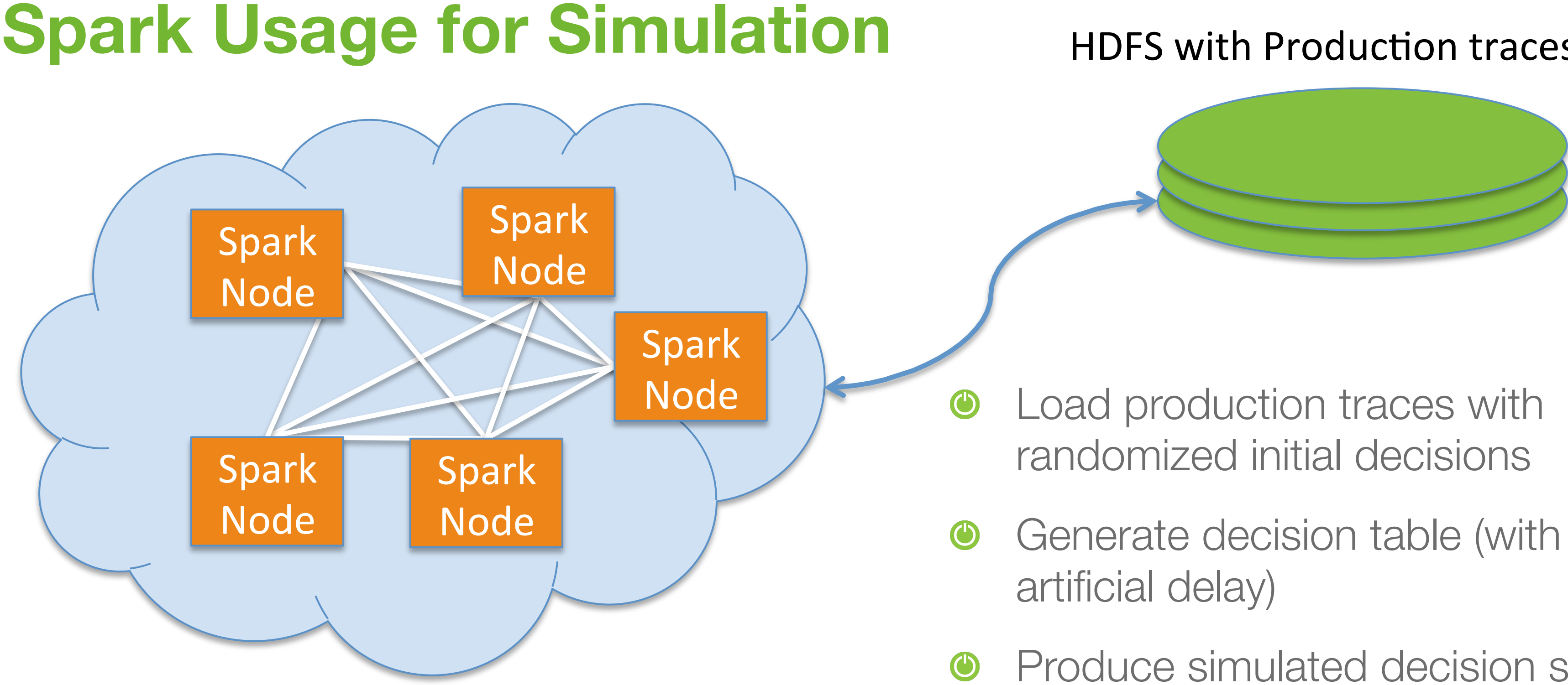
- Can't select a CDN based solely on one metric.
  - Select utility functions that best predict engagement
- Confidence in a decision, or evaluation will depend on how much data we have collected
  - Need to tune time window
  - Select different attributes for separation

CONVIVA

Tuning

- Need to validate algorithm changes quickly
- Simulation of algorithm offline, is essential

Spark Usage for Simulation



- Load production traces with randomized initial decisions
- Generate decision table (with artificial delay)
- Produce simulated decision set
- Evaluate decisions against actual traces to estimate expected quality improvement

CONVIVA

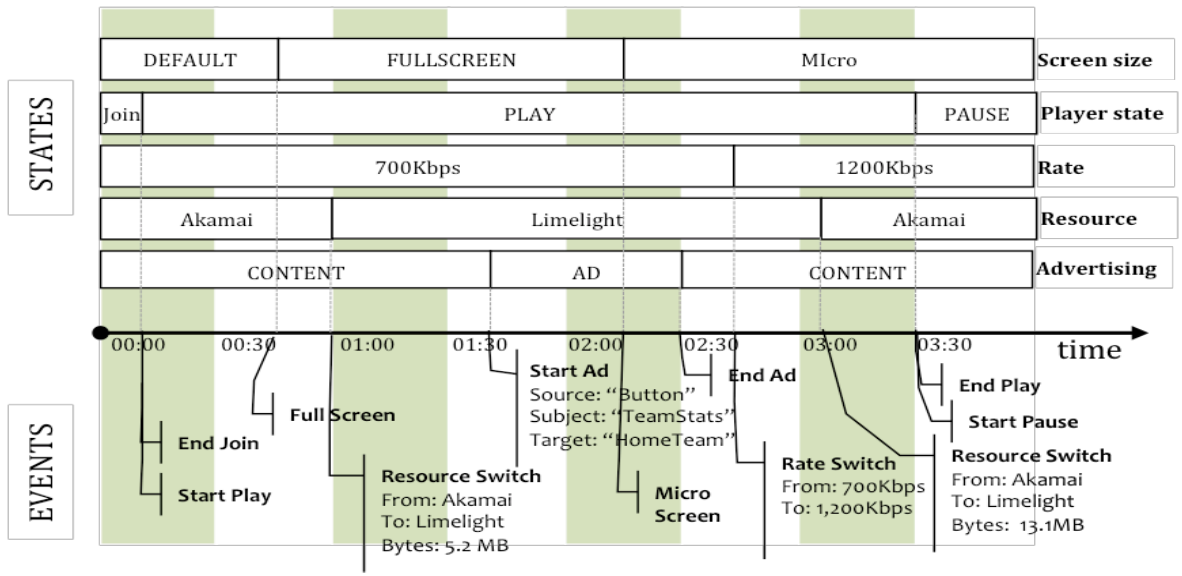
Future of Spark and Conviva

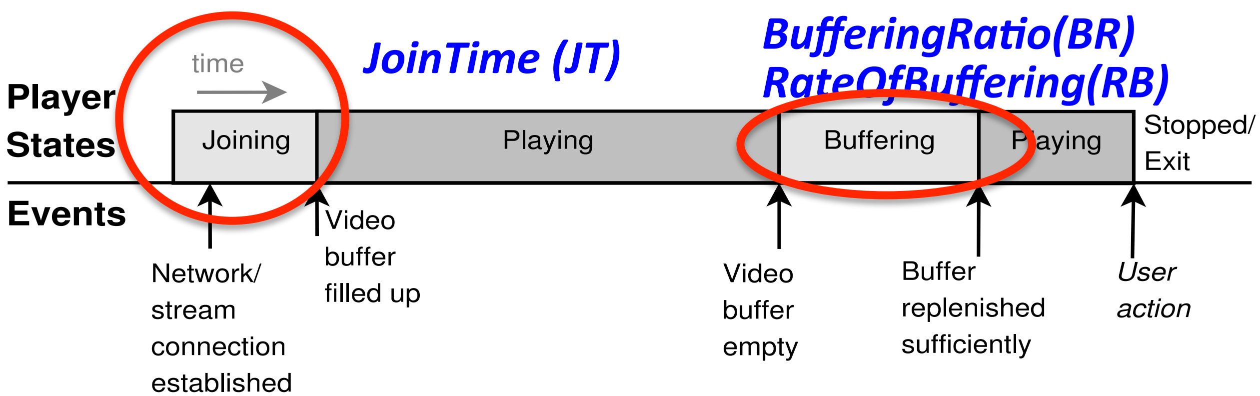
- Leverage spark streaming
- Unify live and historical processing
- Develop platform to build various processing ‘apps’ (e.g. Anomaly Detection, Customer Tailored Reporting)
  - Can share the same data API
  - Will all have consistent input

In Summary

- Spark was able to support our initial requirement of fast fault tolerant performance computation for an on-line decision maker
- New complexities like LP calculation ‘just worked’ in the existing architecture
- Spark has become an essential tool in our software stack

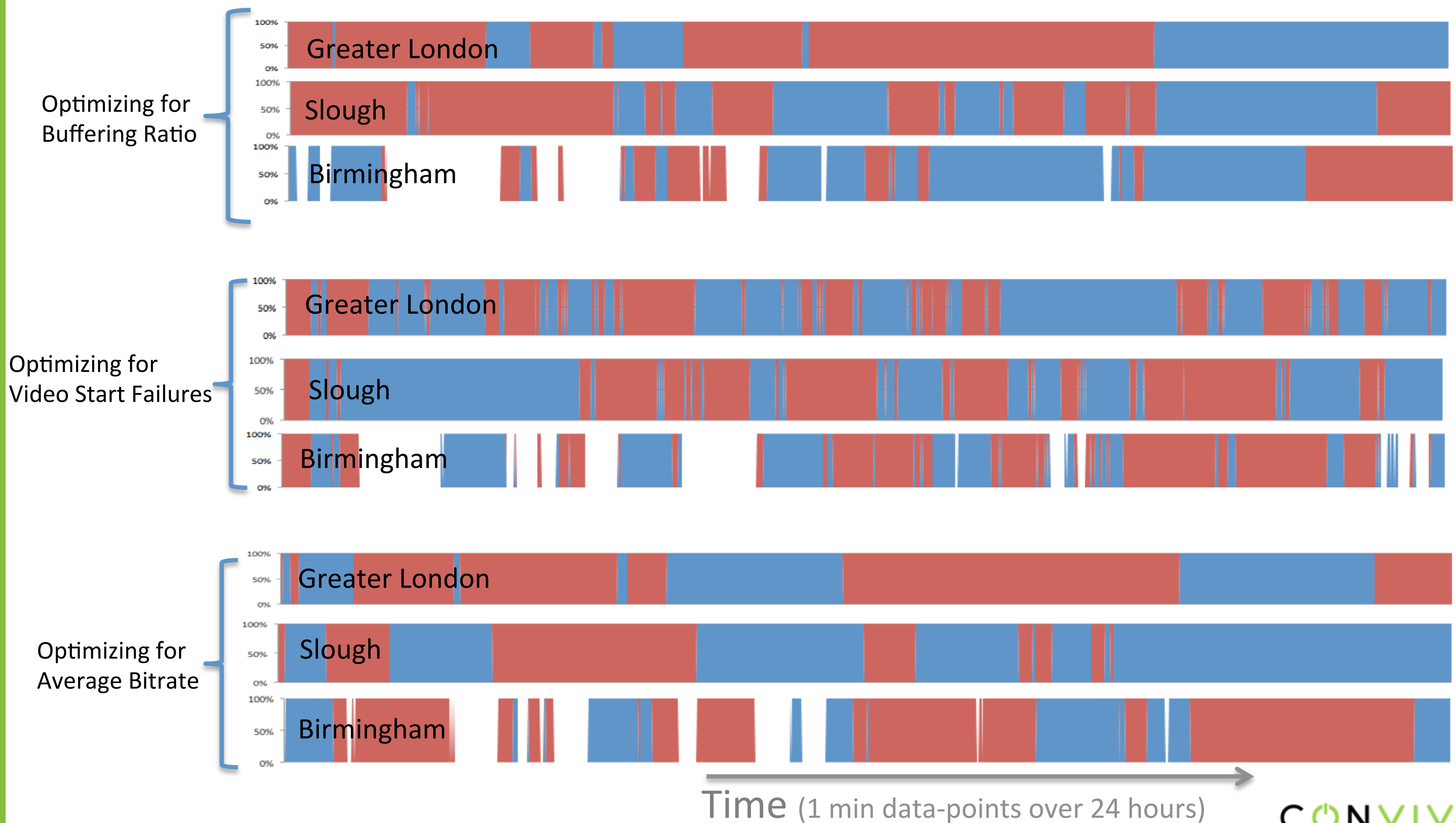
Real-time Measurement from Every Viewer





CONVIVA

An UK ISP: Top Metros by Metric over time



CONVIVA

Measurements of 3 Leading CDN Show Significant Variation Over Geo and ISP