

REDUCING DELIVERY COSTS

by Optimal Multi-CDN Traffic Allocation

Yuriy Reznik, Bo Zhang, Ron Zekarias, Ted Krofssik, Guillem Cabrera*, Stuart Hicks*, Nabajeet Barman*, and Andrew Sinclair**

Brightcove, Inc., (*) Brightcove UK, Ltd., (**) Brightcove Australia, Pty Ltd.

mhv/2023
ACM MILE-HIGH VIDEO
annual workshop

May 7-10, 2023, Denver, CO

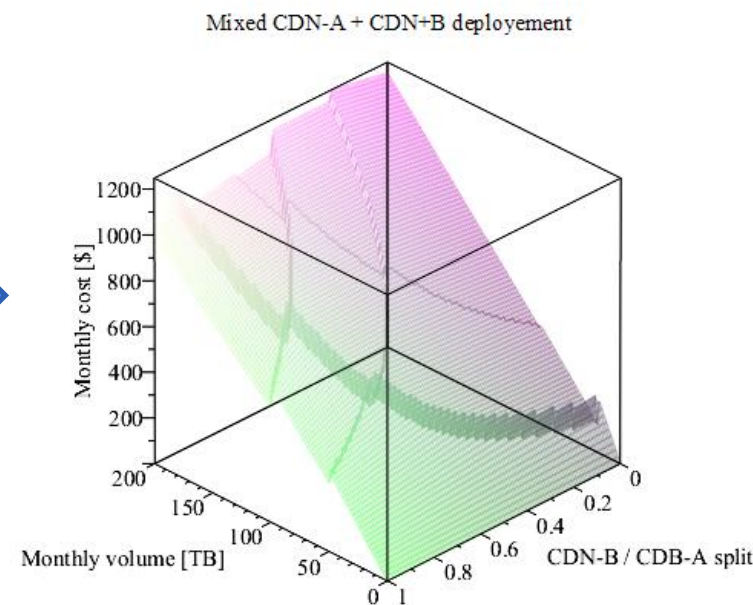
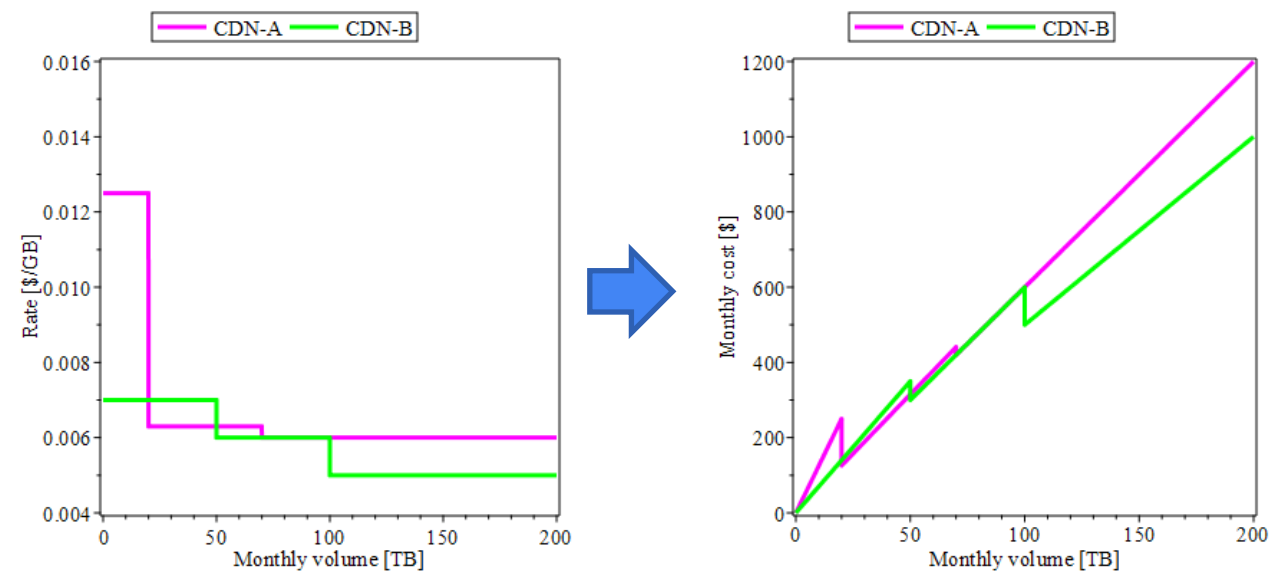
MULTI-CDN DELIVERY

APPLICATIONS / UTILITIES

- ▶ Multi-region delivery
- ▶ Better scale (load balancing)
- ▶ Improved reliability (failover)
- ▶ Improved QOE (QOS/QOE optimizations)

BUT..CAN IT ALSO REDUCE THE COSTS?

- ▶ The intuitive answer is no:
 - > Each CDN comes with a rate ladder
 - > Splitting the volume leads to higher rates



BIG TOPIC

EBU
OPERATING EUROVISION AND EUORADIO

TR 068

CDN ARCHITECTURES
DEMYSTIFIED

AN INSIGHT INTO CDN APPROACHES
EMPLOYED BY EBU MEMBERS

SVTA

Investigating Approaches to Multi-CDN Delivery

When streaming on-demand or live video, issues with a particular CDN can impact the user quality of experience (QoE). If the CDN struggles to fulfil the requests, if the player has access to multiple CDNs, switching to another CDN or using multiple CDNs simultaneously might help maintain a good user experience. Generally speaking, most streaming services use a multi-CDN delivery approach for redundancy and business purposes.

Why Do Streaming Platforms Want to Switch CDNs?

- Performance: A CDN performing poorly in a specific geographic region may compromise QoE. When this happens, it could adversely affect video engagement (how long viewers watch) or even result in churn.
- Capacity: A CDN may sometimes develop specific issues because of competing traffic on its network.
- Traffic controls: A streaming platform may sometimes want to switch because of contractual obligations. For example, a CDN may need to get 50% of traffic, while a CDN may need to get 40%.
- Cost: Streaming platforms may want to switch to CDNs because of lower costs in specific regions or at specific times of the day.

Video Delivery Market 2022–2028
Executive Summary

Open Caching challenges Public, Private CDNs, as WebRTC and Decentralized-CDN tick on

Total Video Traffic - Data and Revenue - Global - 2020-2028

CDN by Service Type 2020 - 2027

streaming media

How to Jump-Start Your Multi-CDN Strategy and Deliver Every Time

MUVI

Everything you Need to Know about Multi CDN Switching

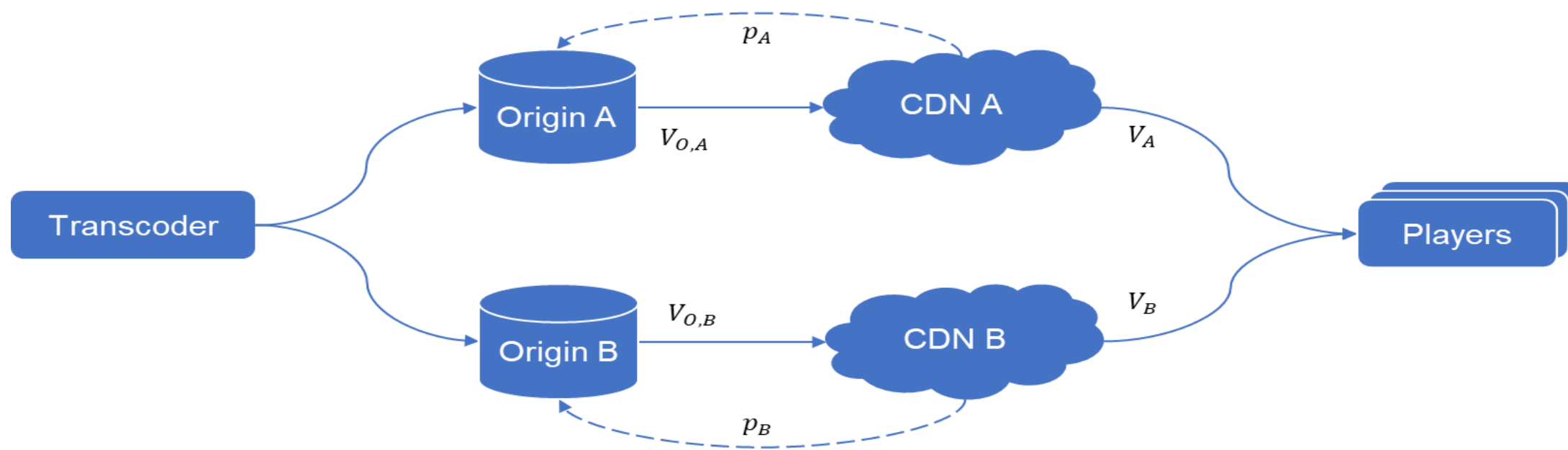
SEAMLESS CDN SWITCHING

CDN Market by Technology, Platform, Application, Service Type, Customer Type, and Industry Verticals 2021 – 2026

MIND COMMERCE

COST MODELS

ARCHITECTURE



VARIABLES

- ▶ Edge volume: V_A, V_B
- ▶ Origin volume: $V_{O,A}, V_{O,B}$
- ▶ Cache miss probabilities: p_A, p_B

RELATIONSHIPS

- ▶ Origin/edge volume:

$$V_{O,A} = V_A \cdot p_A$$

$$V_{O,B} = V_B \cdot p_B$$

COST MODELS

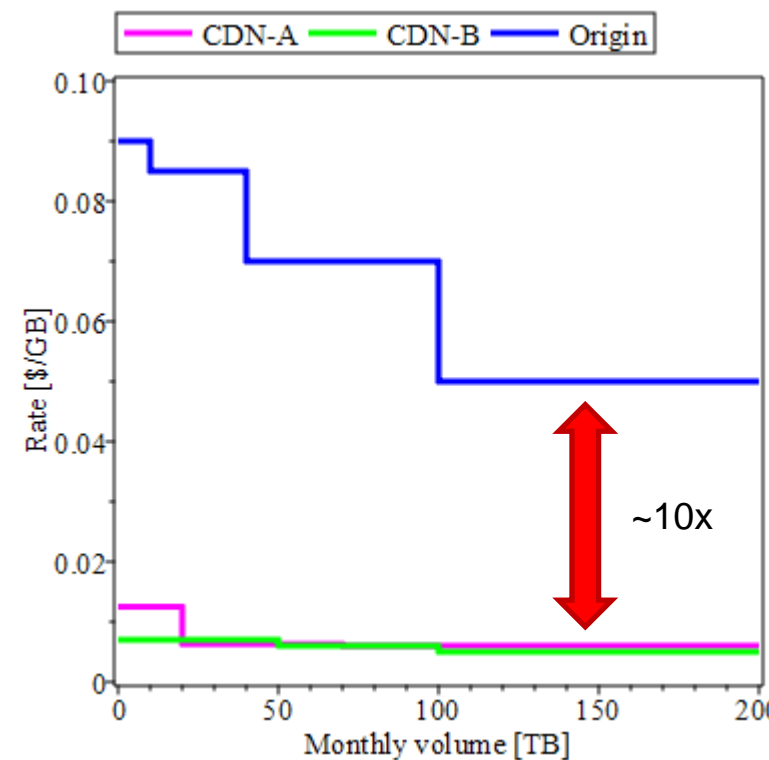
- ▶ CDN costs:

$$C_A = R_A \cdot V_A, \quad R_A = \text{piecewise}(V_A, \dots)$$

$$C_B = R_B \cdot V_B, \quad R_B = \text{piecewise}(V_B, \dots)$$
- ▶ CDN commits add extra constraints
- ▶ Origin costs:

$$C_{O,A} = R_{O,A} \cdot V_{O,A}, \quad R_{O,A} = \text{piecewise}(V_{O,A}, \dots)$$

$$C_{O,B} = R_{O,B} \cdot V_{O,B}, \quad R_{O,B} = \text{piecewise}(V_{O,B}, \dots)$$
- ▶ Generally, origins are more expensive



FULL COSTS

- ▶ Origin + CDN costs:

$$C_{\Sigma,A} = C_{O,A} + C_A = V_A(p_A \cdot R_{O,A} + R_A)$$

$$C_{\Sigma,B} = C_{O,B} + C_B = V_B(p_B \cdot R_{O,B} + R_B)$$
- ▶ Effective rates:

$$R_{\Sigma,A} = p_A \cdot R_{O,A} + R_A$$

$$R_{\Sigma,B} = p_B \cdot R_{O,B} + R_B$$

LESS EXPENSIVE PATHWAY

COMPARING THE COSTS

- ▶ Assume that origin costs are the same:

$$R_{O,A} = R_{O,B} = R_O$$

- ▶ Effective rates along each pathway:

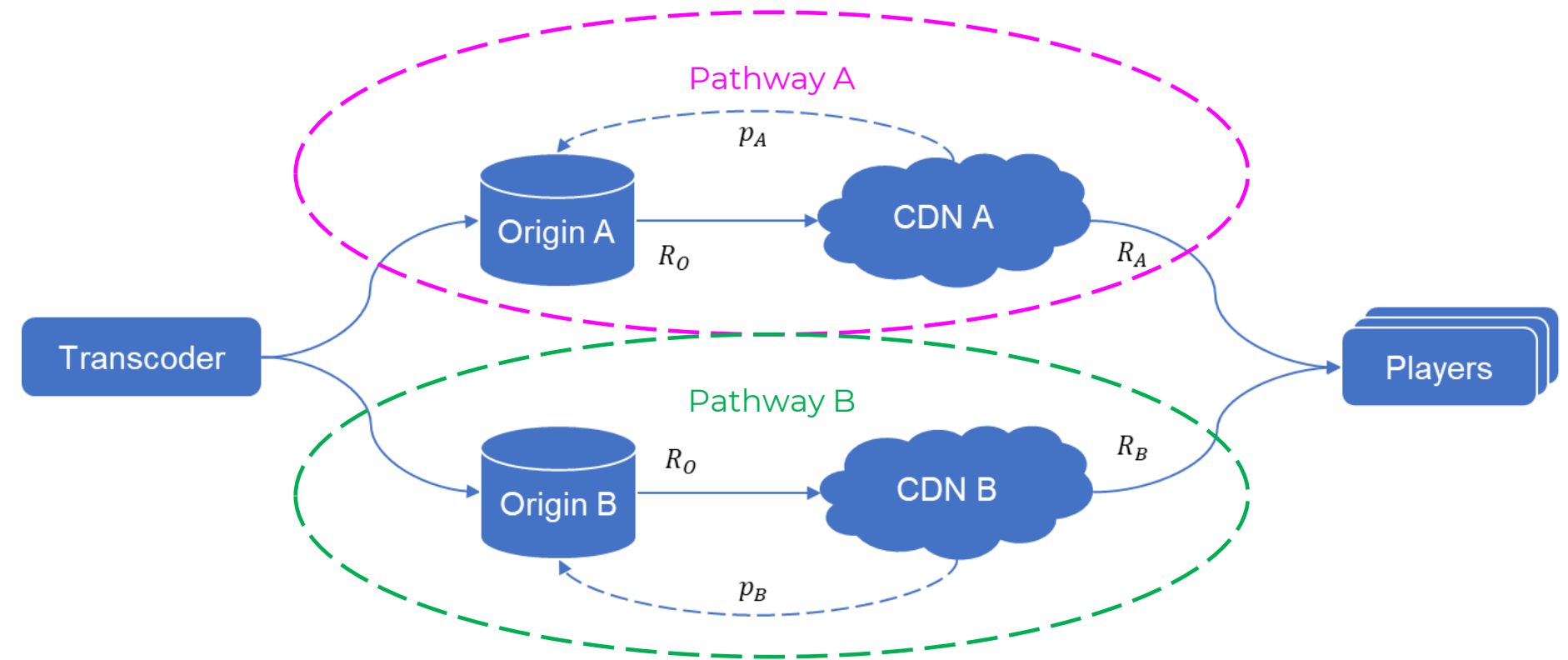
$$R_{\Sigma,A} = p_A \cdot R_O + R_A$$

$$R_{\Sigma,B} = p_B \cdot R_O + R_B$$

- ▶ Pathway A is less expensive if:

$$R_{\Sigma,A} < R_{\Sigma,B} \Rightarrow p_A \cdot R_O + R_A < p_B \cdot R_O + R_B$$

$$\Rightarrow p_A - p_B < \frac{R_B - R_A}{R_O} = \xi$$



EXAMPLES

SYNOPSIS	R_O	R_A	R_B	ξ	p_A	p_B	$p_A - p_B$	LESS EXPENSIVE PATHWAY
CDN A is cheaper & better in cache performance	0.02	0.002	0.0025	0.025	0.07	0.1	$-0.03 < \xi$	A
CDN A is cheaper & worse in cache performance	0.02	0.002	0.0025	0.025	0.1	0.07	$0.03 > \xi$	B
CDN A is more expensive & better in cache performance	0.02	0.0025	0.002	-0.025	0.07	0.1	$-0.03 < \xi$	A
CDN A is more expensive & worse in cache performance	0.02	0.0025	0.002	-0.025	0.1	0.07	$0.03 > \xi$	B

NB: Cache performance has a major impact on overall costs and choice of best pathway.

IMPACT OF CONTENT POPULARITY

FUNDAMENTAL RELATIONSHIP

- ▶ If content is popular, it gets cached with higher probability
- ▶ But.... such relationships may vary across CDNs, regions, and under different load
- ▶ To describe them we may use parametric models, e.g.:

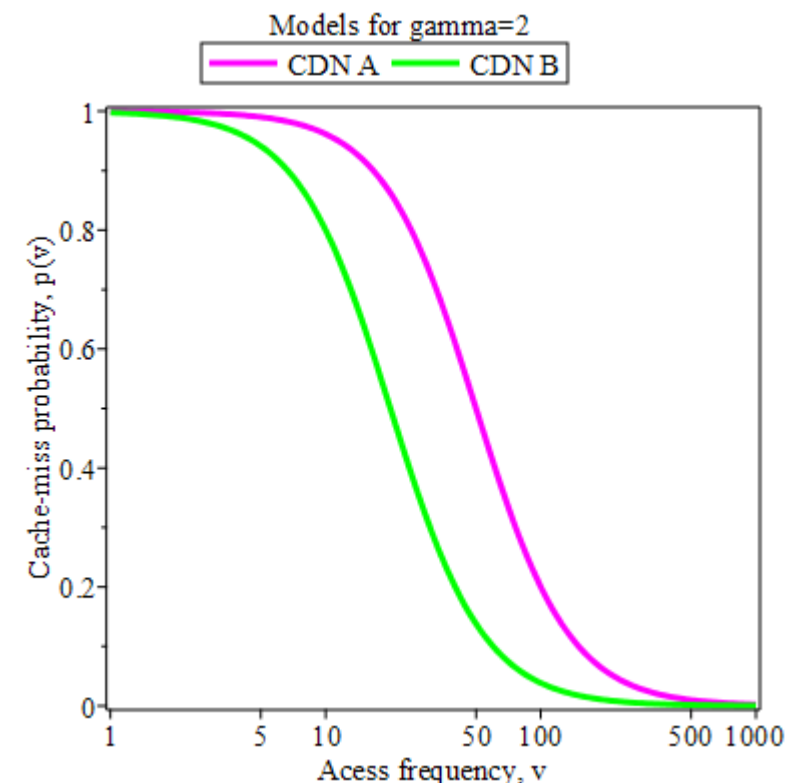
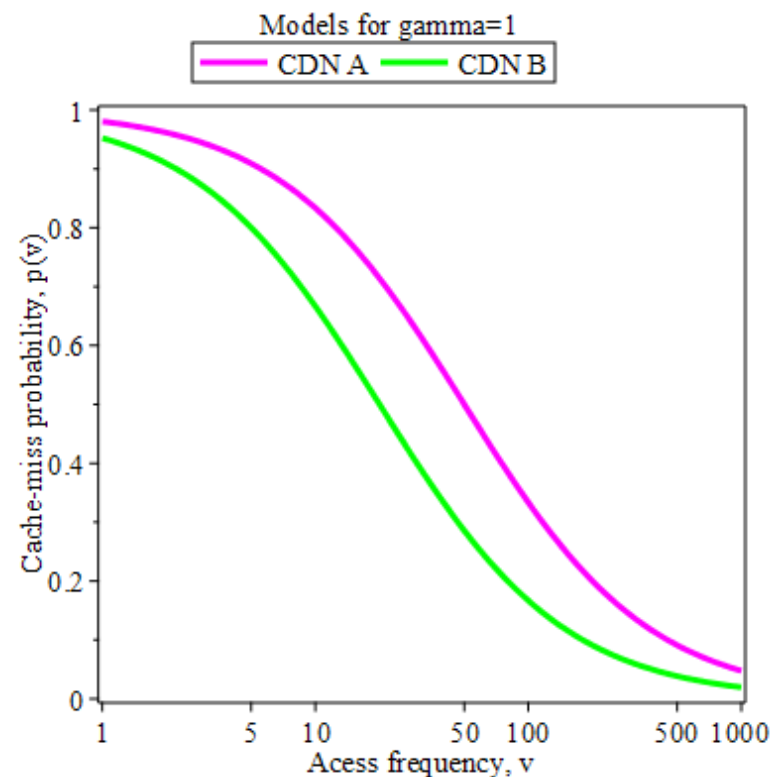
$$p_A(v) = \frac{1}{1 + (v/v_A)^\gamma}, \quad p_B(v) = \frac{1}{1 + (v/v_B)^\gamma};$$

where

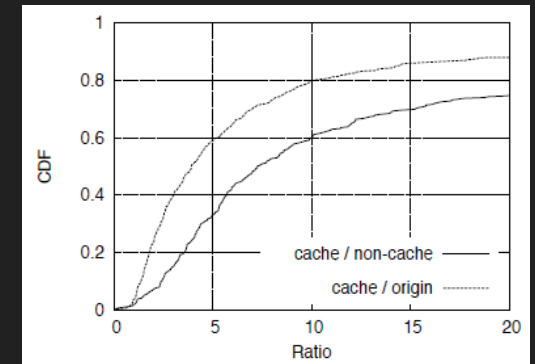
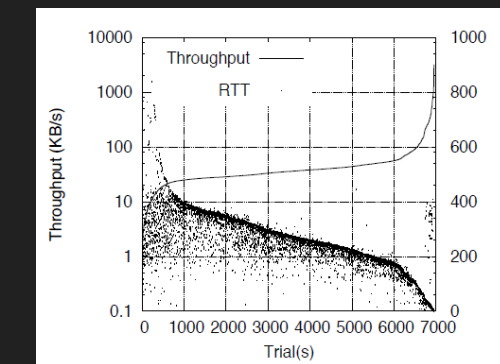
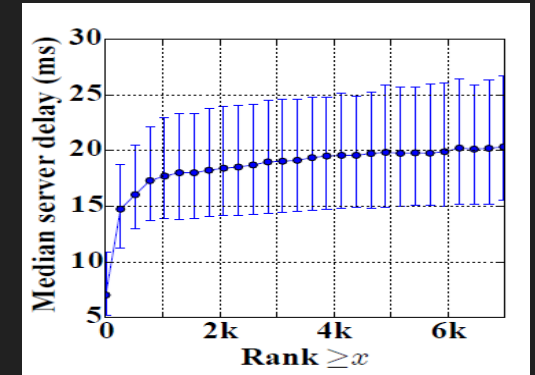
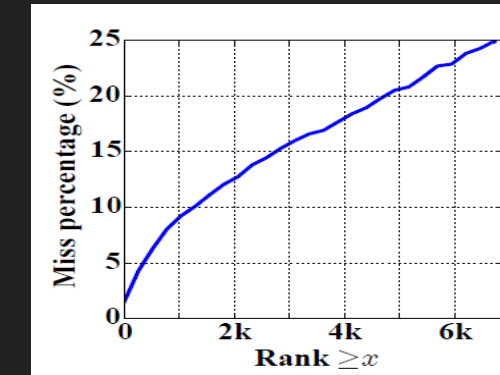
- v – content access frequency (e.g. requests/day)
- $p_A(v), p_B(v)$ – cache miss probabilities of CDN A and CDN B, respectively
- v_A, v_B – CDN-specific model parameters
- γ – model shape parameter

EXAMPLES

- ▶ Assume that $v_A = 50, v_B = 20$
- ▶ Model plots for $\gamma = 1, 2$:



Related publications / studies



P. Franaszek and T. Wagner, "Some distribution-free aspects of paging algorithm performance," JACM, 21(1), 1974, pp.31-39.

P. R. Jelenkovic, "Asymptotic approximation of the move-to-front search cost distribution and least-recently-used caching fault probabilities," Ann. Appl. Probab., 9 (2), 1999, pp. 430–464.

S. Triukose, Z. Wen, and M. Rabinovich, "Measuring a Commercial Content Delivery Network," ACM WWW, 2011.

M. Ghasemi, P. Kanuparth, A. Mansy, T. Benson, and J. Rexford, "Performance characterization of a commercial video streaming service," ACM ICM, 2016.

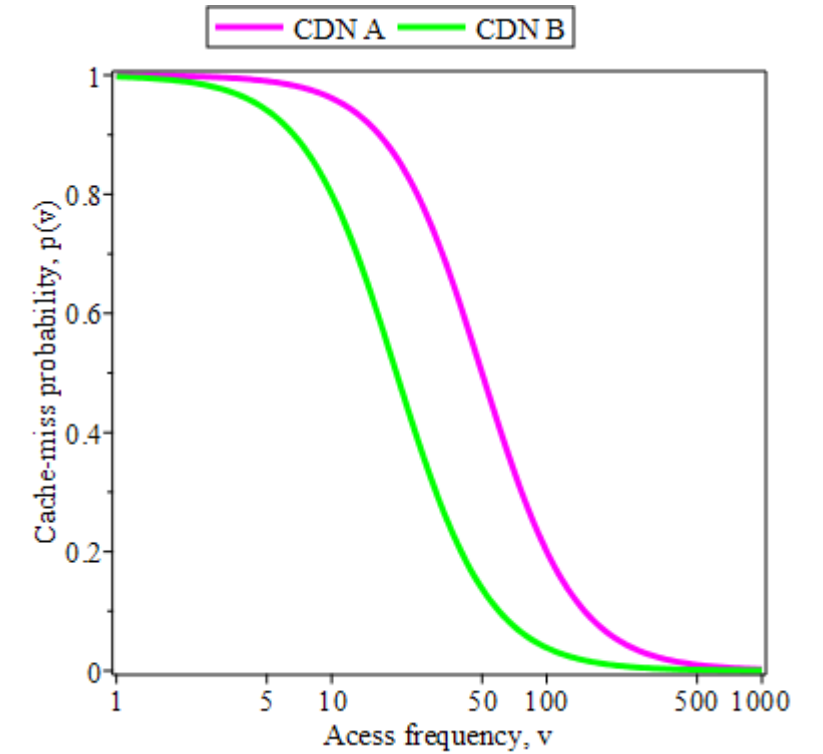
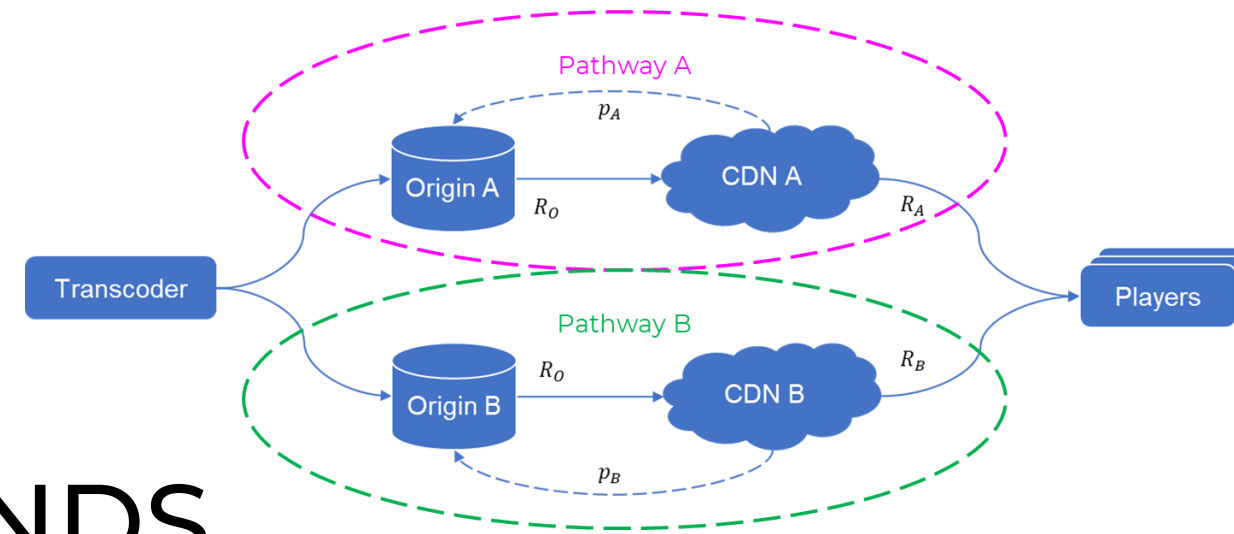
Y. Reznik, T. Teixeira, and R. Peck, "On multiple media representations and CDN performance," ACM MHV, 2022.

FINDING LESS EXPENSIVE PATHWAY

INITIAL SOLUTION

- ▶ For fixed p_A, p_B , we already established that:

$$R_{\Sigma,A} < R_{\Sigma,B} \Rightarrow p_A - p_B < \frac{R_B - R_A}{R_O} = \xi$$



ACCESS FREQUENCY BOUNDS

- ▶ Models for cache miss probabilities:

$$p_A(v) = \frac{1}{1 + (v/v_A)^\gamma}, \quad p_B(v) = \frac{1}{1 + (v/v_B)^\gamma}$$

- ▶ Translation to bounds for access frequency v :

$$R_{\Sigma,A} < R_{\Sigma,B} \Rightarrow v \in \begin{cases} [0, \infty) & \text{if } v_A < v_B \text{ and } R_A < R_B \\ (v_1^*, v_2^*) & \text{if } v_A < v_B \text{ and } R_A > R_B \\ [0, v_1^*) \cup (v_2^*, \infty) & \text{if } v_A > v_B \text{ and } R_A < R_B \\ \emptyset & \text{if } v_A > v_B \text{ and } R_A > R_B \end{cases}$$

- ▶ Where $v_1^* < v_2^*$ are the real positive roots of

$$\frac{1}{1 + (v/v_A)^\gamma} - \frac{1}{1 + (v/v_B)^\gamma} = \xi$$

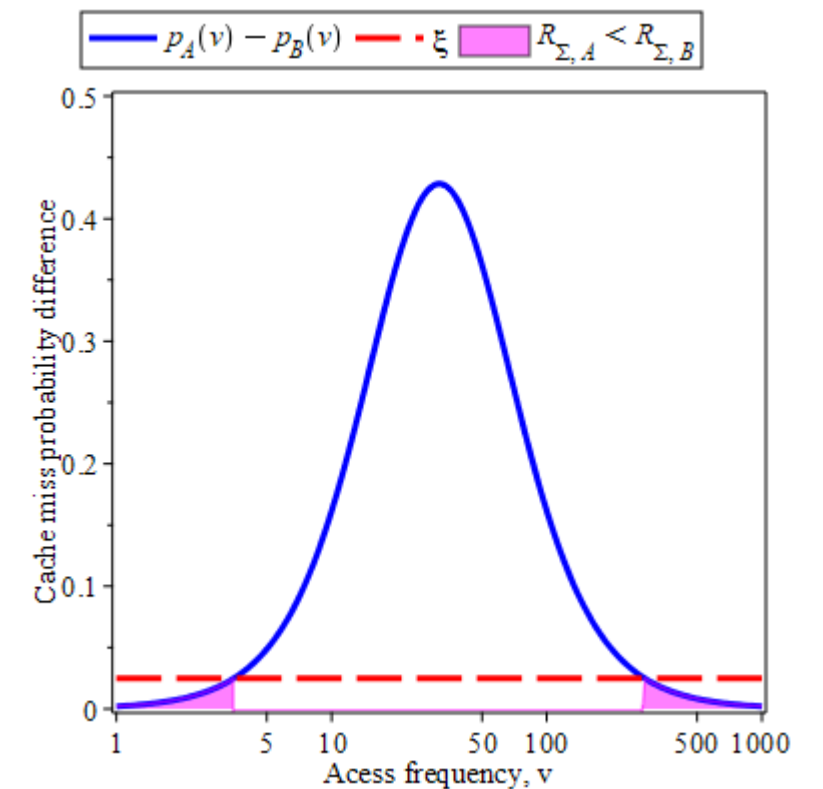
- ▶ E.g., for $\gamma = 2$:

$$v_1^* = \frac{1}{\sqrt{2\xi}} \sqrt{v_A^2 - v_B^2 - \xi(v_A^2 + v_B^2) - \sqrt{(v_A^2 - v_B^2)(v_A - v_B - \xi(v_A + v_B))(v_A + v_B - \xi(v_A - v_B))}}$$

$$v_2^* = \frac{1}{\sqrt{2\xi}} \sqrt{v_A^2 - v_B^2 - \xi(v_A^2 + v_B^2) + \sqrt{(v_A^2 - v_B^2)(v_A - v_B - \xi(v_A + v_B))(v_A + v_B - \xi(v_A - v_B))}}$$

EXAMPLE

- ▶ CDN A is cheaper: $R_A < R_B$, $\xi = 0.025$
- ▶ But worse as a cache: $v_A = 50, v_B = 20$
- ▶ Roots: $v_1^* \approx 3.51$ and $v_2^* \approx 284.76$
- ▶ The solution: $v \in [0, v_1^*) \cup (v_2^*, \infty)$
- ▶ NB: using pathway A in this case makes sense only for high access or long tail content!



VOD: BEST PER-ASSET CDN ASSIGNMENT

CONSIDER A LARGE CATALOG

- ▶ Videos are ordered according to access frequencies
- ▶ Follow Zeta distribution:

$$u(i) = \zeta(\alpha)^{-1} i^{-\alpha}$$

where α is a shape parameter, and $\zeta(\alpha)$ is the Riemann's Zeta function, i is an asset index.

BEST PER-ASSET CDN ASSIGNMENT

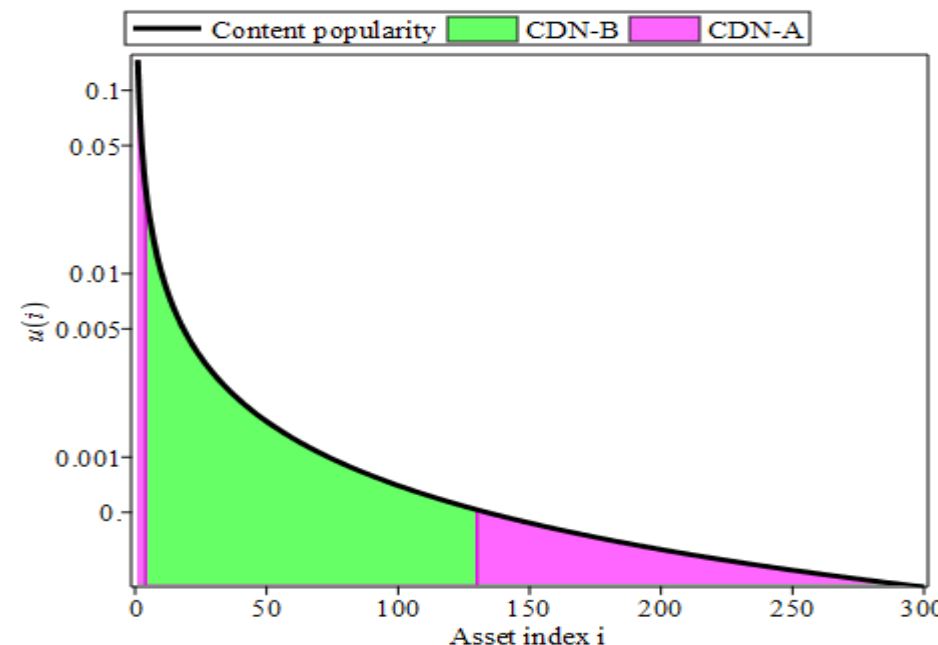
- ▶ For given CDN prices R_A, R_B and cache miss models $p_A(v), p_B(v)$, we can show that:

$$R_{\Sigma,A} < R_{\Sigma,B} \Rightarrow i \in \begin{cases} [1, \infty) & \text{if } v_A < v_B \text{ and } R_A < R_B \\ (i_1^*, i_2^*) & \text{if } v_A < v_B \text{ and } R_A > R_B \\ [1, i_1^*) \cup (i_2^*, \infty) & \text{if } v_A > v_B \text{ and } R_A < R_B \\ \emptyset & \text{if } v_A > v_B \text{ and } R_A > R_B \end{cases}$$

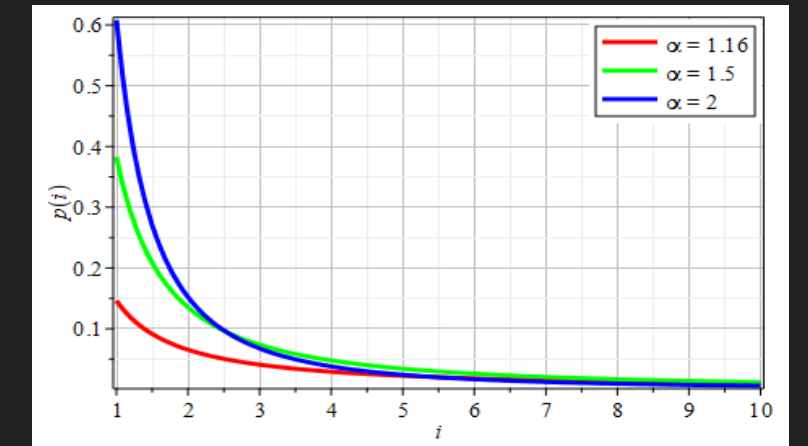
where $i_1^* = (v_2^*/C_\alpha)^{-1/\alpha}$, $i_2^* = (v_1^*/C_\alpha)^{-1/\alpha}$ are boundary points, C_α - normalization constant

EXAMPLE

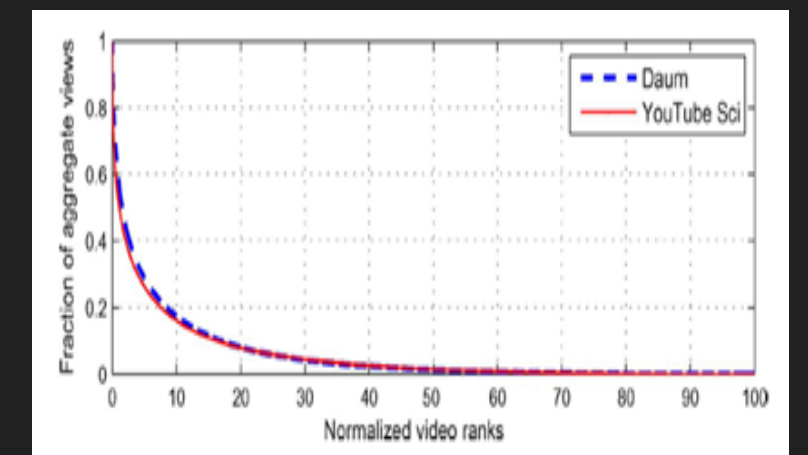
- ▶ CDN A is cheaper: $R_A < R_B$, $\xi = 0.025$
- ▶ Worse as cache: $v_A = 50, v_B = 20, \gamma = 2$
- ▶ Roots: $v_1^* \approx 3.51$ and $v_2^* \approx 284.76$
- ▶ Content distribution: $\alpha = 1.16, C_\alpha = 1000$
- ▶ Boundary points: $i_1^* \approx 3, i_2^* \approx 130$
- ▶ Solution for CDN-A: $i \in [1, i_1^*) \cup (i_2^*, \infty)$



Zeta distribution



Ranks of videos in YouTube



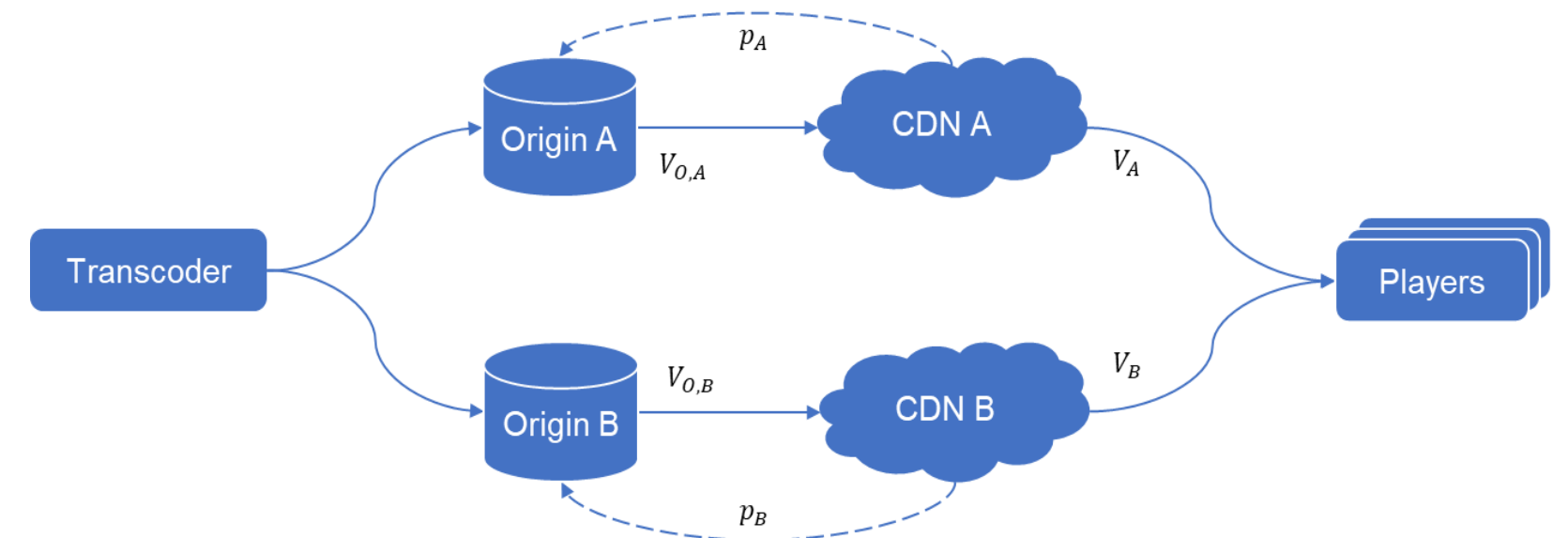
M. Cha, et al, "Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems," IEEE/ACM Trans. Networks, vol. 17, 2009, pp 1357-1370.

N. Kamiyama and M. Murata, "Reproducing Popularity Distribution of YouTube Videos," in IEEE Transactions on Network and Service Management, vol. 16, no. 3, 2019, pp. 1100-1112.

VOD: COST OPTIMIZATION PROBLEM

GIVEN

- ▶ $R_A(V), R_B(V)$ – price/rate ladders for CDNs A and B
- ▶ $V_{A,\min}, V_{B,\min}$ – minimum volume commits for each CDN
- ▶ $R_{O,A}(V), R_{O,B}(V)$ – rate ladders for origins A and B
- ▶ $p_A(v, V), p_B(v, V)$ – cache miss models for CDNs A and B
- ▶ $u(i), i \in [1, N]$ – content popularity distribution across catalog
- ▶ $V_\Sigma = V_A + V_B$ – total volume delivered by the system



FIND

- ▶ i_A^*, i_B^* : $i_A^* \cup i_B^* = [1, N]$ – subsets of catalog items routed to CDN A and B, respectively

SUCH THAT

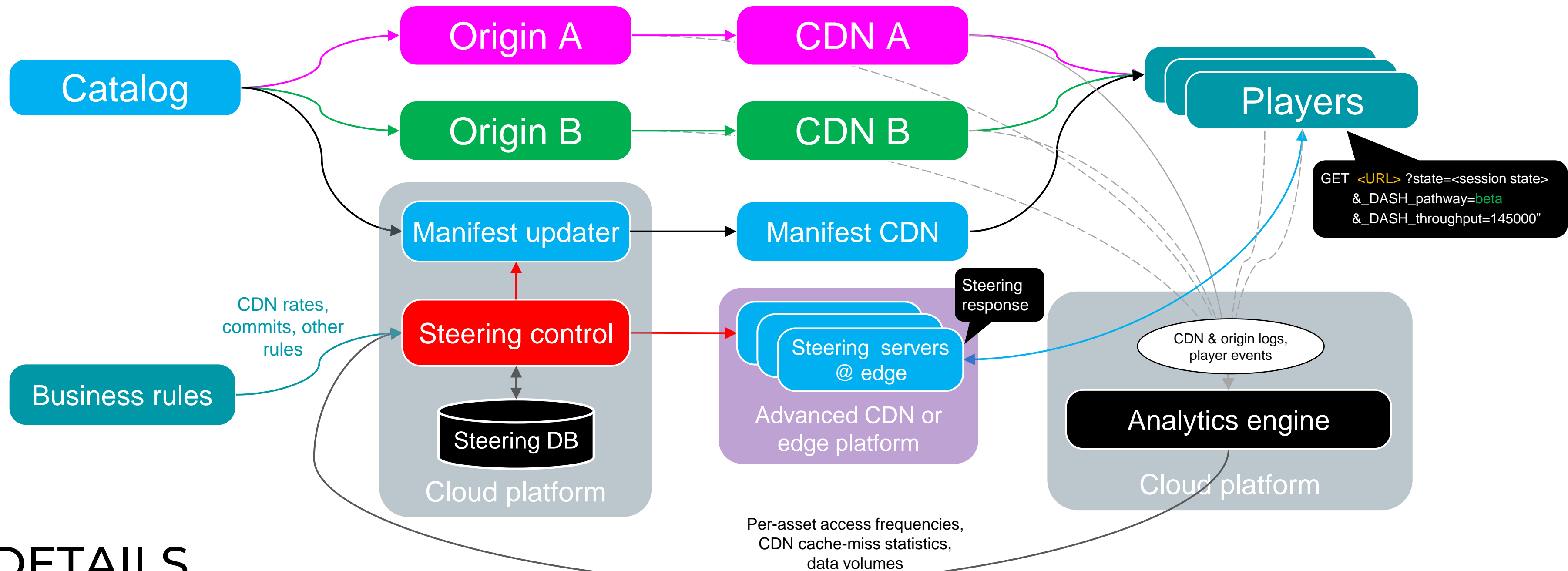
$$C_A(i_A^*) + C_B(i_B^*) = \min_{\substack{i_A, i_B: i_A \cup i_B = [1, N] \\ V_A(i_A) \geq V_{A,\min} \\ V_B(i_B) \geq V_{B,\min}}} C_A(i_A) + C_B(i_B)$$

WHERE

- ▶ $V_A(i_A) = \sum_{i \in i_A} V_\Sigma u(i)$, $V_B(i_B) = \sum_{i \in i_B} V_\Sigma u(i)$ – edge volumes delivered by CDN A and B, respectively
- ▶ $V_{O,A}(i_A) = \sum_{i \in i_A} V_\Sigma u(i) \cdot p_A(V_\Sigma u(i), V_A(i_A))$, $V_{O,B}(i_B) = \sum_{i \in i_B} V_\Sigma u(i) \cdot p_B(V_\Sigma u(i), V_B(i_B))$ – volumes processed by each origin server
- ▶ $C_A(i_A) = V_{O,A}(i_A) \cdot R(V_{O,A}(i_A)) + V_A(i_A) \cdot R_A(V_A(i_A))$, $C_B(i_B) = V_{O,B}(i_B) \cdot R(V_{O,B}(i_B)) + V_B(i_B) \cdot R_B(V_B(i_B))$ – total costs along each pathway

IMPLEMENTATION

USING HLS/DASH CONTENT STEERING FRAMEWORK



DETAILS

- ▶ All cost-related optimization decisions are made by **steering control** logic
- ▶ Manifest updaters and edge-based steering servers are used to enforce these decisions
- ▶ Moving steering servers to edge reduces the costs of multi-CDN management

CONCLUSIONS

MULTI-CDN DELIVERY IS REAL

- ▶ Brings numerous advantages and utilities (scale, reliability, QOE/QOS)
- ▶ With a diverse set of CDNs and popularity of content – **it can be used to reduce costs of delivery !**
- ▶ This is rather unexpected, but very important utility!

CHALLENGES

- ▶ Requires solving a rather intricate optimization problem
- ▶ Requires minimum overhead of CDN traffic steering solution
- ▶ Requires low-cost real-time analytics

VARIATIONS AND FUTURE WORK

- ▶ More complex architectures (cascaded CDNs, etc.)
- ▶ Multi-regional streaming (optimal multi-regional traffic allocation)
- ▶ Live and mixed live + VOD delivery cases



**THANK
YOU**