brightcove

CONTEXT-AWARE ENCODING AND 5G

DASH-IF WORKSHOP ON MEDIA STREAMING MEETS 5G, DECEMBER 9-10, 2019, PORTLAND, OR

YURIY REZNIK | YREZNIK@BRIGHTCOVE.COM

OUTLINE

- Brief history of context-aware encoding & variants
- How it works
- Types of CAE technologies today
- CAE and standards
  - CAE technologies that are fully compatible with existing standards & players
  - CAE technologies that may need extensions
- Discussion
  - CAE and 5G
    - is there any overlap?
    - what type of information from 5G network layer could be useful for CAE
  - Shall there be extension of a standard?
    - E.g. for per-scene ladder signaling?

BRIEF HISTORY OF CAE

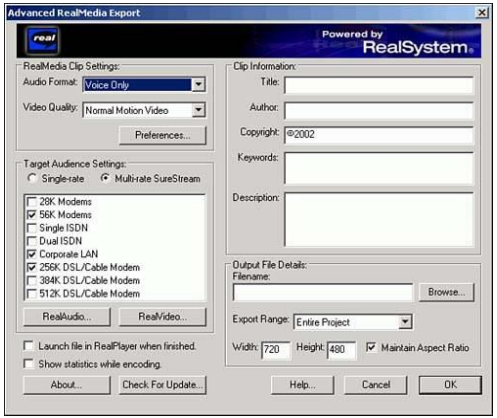
- Early 1990s: H.261, H.263, MPEG-1/2
  - “fixed QP” regime – the grandfather of everything “Constant Quality”
- Late 1990s-early 2000s: RealVideo 8-10
  - “RMVB” – heavy VBR encoding regime, optimized for downloads, still in use in Asia
- Early 2010s:
  - British Telecom “Quality-driven streaming” (Mike Nilsson, et al, June 2012)
  - InterDigital “Quality-based streaming” proposal to MPEG-DASH (Y. Reznik et al, m25996, July 2012)
  - Intel labs “Quality-aware streaming” (Yiting Liao, et al, 2013)
  - “Capped-CRF” – approaches – multiple sources, 2013+
  - Beamr “Optimizer” – second pass encoding with adjusted targets, 2013+
  - MediaMelon “QBR streaming”, 2014
- Late 2010s:
  - Netflix “Per-Title Encoding” blog post, Dec. 2015 – ladder of resolutions and rates according to content
  - Brightcove “Context-Aware Encoding”, Oct. 2016 – ladder design as end-to-end optimization problem
  - Netflix “per-scene encoding”, 2018 – same as per-title, but on scene basis
  - Content- and context-aware solutions from Harmonic, Elemental, Ateme, Bitmovin, EpicLabs, Mux, etc.

STATIC ABR ENCODING PROFILES

- Define sets of encoding parameters for each rendition:
  - Resolutions, Bitrates, Codec constraints, etc.
  - Same for all content, networks, user devices & usage patterns, etc.

- Some examples of ABR profiles used in practice:

RealVideo (1998):



Apple HLS guidelines (2018):

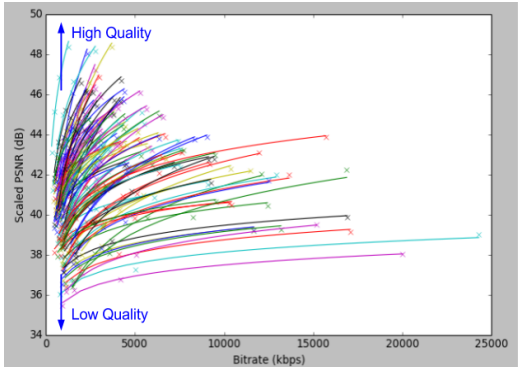
HEVC/H.265	H.264/AVC	Resolution	Frame rate
145	145	416 x 234	≤ 30 fps
350	365	480 x 270	≤ 30 fps
660	730	640 x 360	≤ 30 fps
990	1100	768 x 432	≤ 30 fps
1700	2000	960 x 540	same as source
2400	3000	1280 x 720	same as source
3200	4500	same as source	same as source
4500	6000	same as source	same as source
5800	7800	same as source	same as source

Brightcove VideoCloud (2013-2016):

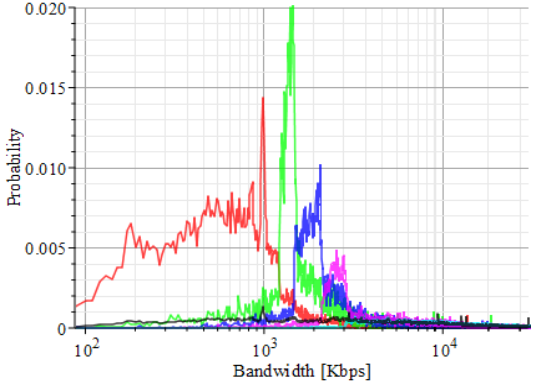
video bitrate	decoder bitrate cap	decoder buffer size	max frame rate	width	height	h264 profile
450	771	1028	30	480	270	baseline
700	1194	1592	30	640	360	baseline
900	1494	1992	30	640	360	main
1200	1944	2592	30	960	540	main
1700	2742	3656	30	960	540	main
2500	3942	5256	30	1280	720	main
3500	5442	7256	30	1920	1080	high
3800	6192	8256	30	1920	1080	high

WHY STATIC ABR PROFILES ARE BAD?

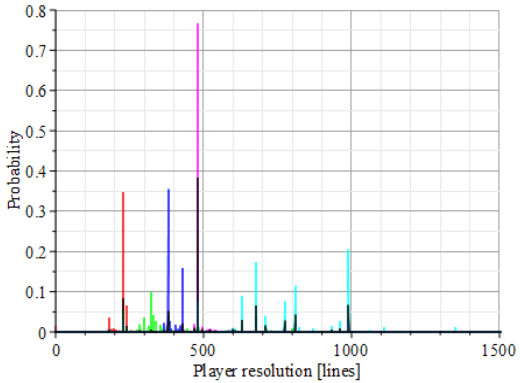
- Static encoding profiles are not accounting for:
  - differences in video complexity:
  - differences in networks:
  - differences in devices & user preferences:



Source: Netflix, 2015



Source: Brightcove VideoCloud analytics, 2019

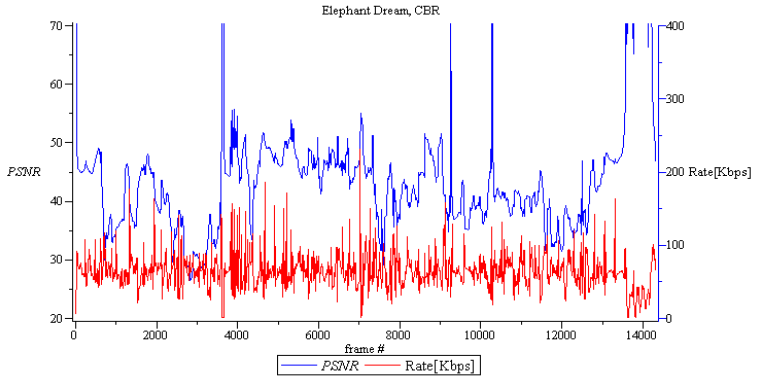


Source: Brightcove VideoCloud analytics, 2019

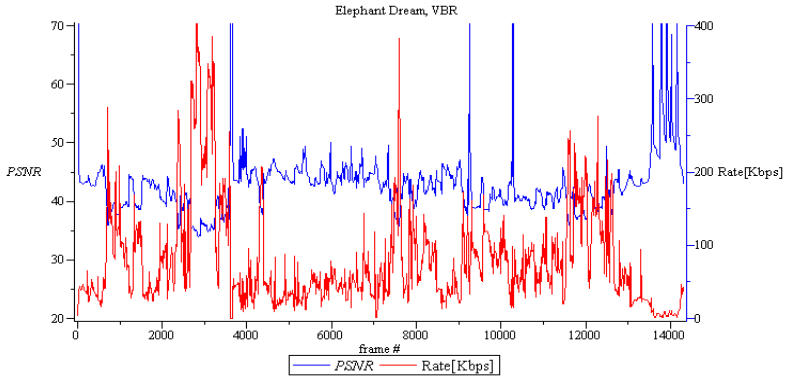
- **A better approach is to design encoding profiles dynamically,** accounting for characteristics of
  - content → content-aware encoding (aka per-title encoding)
  - network → network-aware encoding
  - full context (content + network + user statistics) → **context-aware encoding**

CAE APPROACHES: “CONTENT-AWARE VBR ENCODING”

- Basically, most encoders can be configure to operate either in
  - **“CBR” mode** => reduces variation of bitrates:



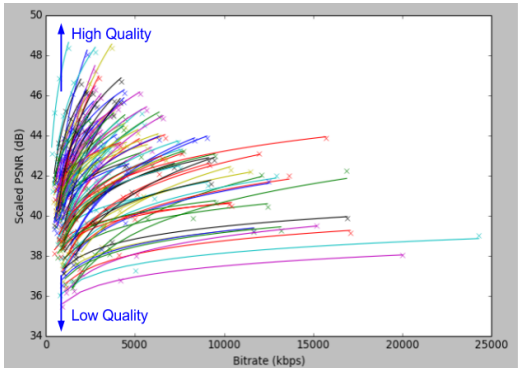
  - **“VBR” mode** => reduces variation of quality:

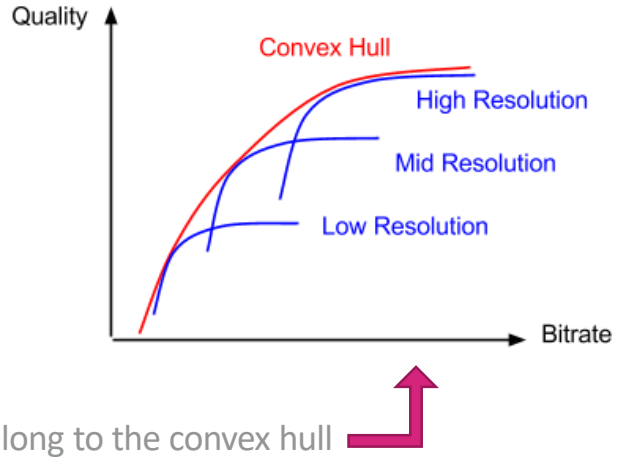


- CBR is required for cable & broadcast (c.f. SCTE 128)
- VBR (with some reasonable constraints) is working reasonably well for OTT
  - Apple HLS constraints (2018):
    - Live: max bitrate < 110% of target
    - VOD: max bitrate < 200% of target (in practice it is better to limit it to about 150%)
- Reasons for constraints: minimize client’s mis-predictions, likelihood of buffering, issues with analytics, etc.

CAE APPROACHES: “PER-TITLE” AND “PER-SCENE ENCODING”

- Primary idea: design ABR encoding profiles individually for each video sequence (or scene within a sequence)



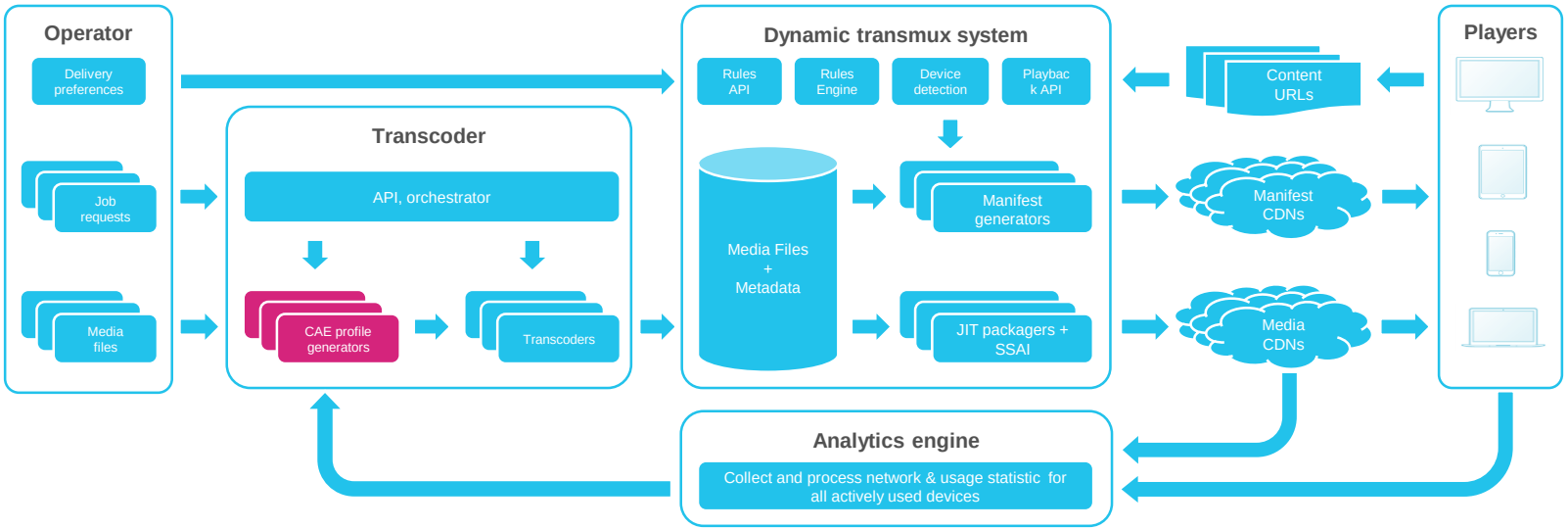


- Secondary idea: place ladder points such that they belong to the convex hull

- Notes:
  - Netflix “convex hull” argument provides a method for finding best resolutions for any given target bitrate, but
    - it does not, say **how** such bitrates should be placed, or **how many** of them are needed!
    - **it constrains the problem, but it does not show how to solve it exactly!**

CAE APPROACHES: “CONTEXT-AWARE ENCODING”

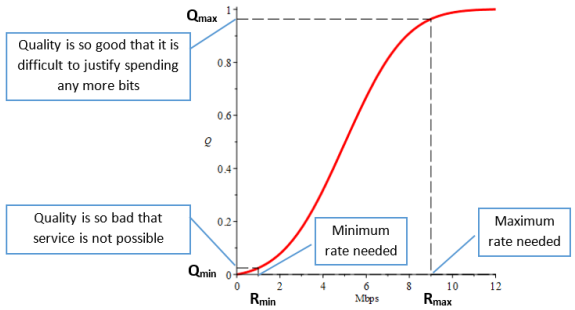
- Example deployment architecture:



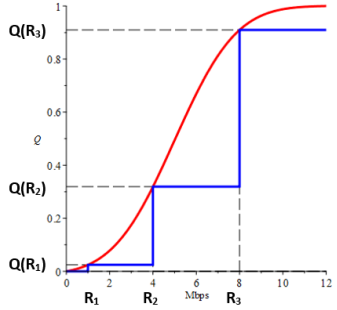
- Context Aware Encoding (CAE) is basically an
  - ABR encoding profile generator that considers:
    - properties of content and
    - properties of networks and devices used to receive content

CONTEXT-AWARE ENCODING: THE PRINCIPLE

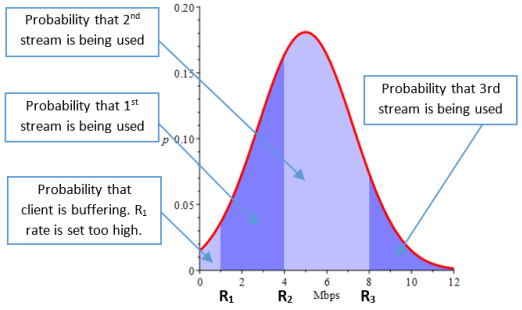
- Quality-rate function $Q(R)$:



Quality delivered by streaming client:



Probabilities of loading of each stream:



- Average quality for a given ladder of rates $R_1, \ldots, R_n$, quality-rate function $Q(R)$, and network density $p(R)$:

$$\overline{Q}(R_1, \ldots, R_n, p) = Q(R_1) \int_{R_1}^{R_2} p(R)dR + Q(R_2) \int_{R_2}^{R_3} p(R)dR + \ \ldots + Q(R_n) \int_{R_n}^{R_{\max}} p(R)dR$$

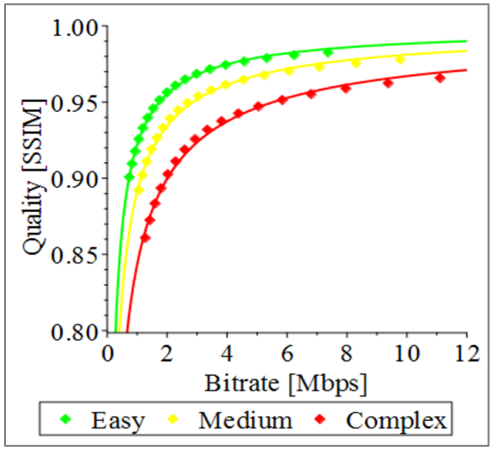
- A **quality-optimal ladder** is a set of rates $\hat{R}_1, \ldots, \hat{R}_n$, such that:

$$\overline{Q}(\hat{R}_1, \ldots, \hat{R}_n, p) = \max_{\substack{R_{\min} < R_1 \leq \cdots \leq R_n < R_{\max} \\ R_1 \leq R_{1,\max}}} \overline{Q}(R_1, \ldots, R_n, p)$$

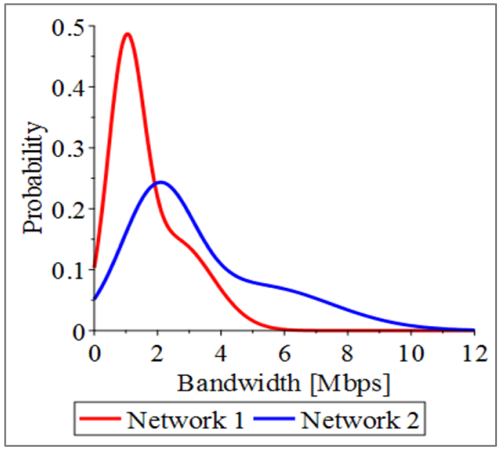
CONTEXT-AWARE ENCODING: EXAMPLE INPUTS

- Content:

Resolution=720p25
Codec=H.264/Main
Quality metric=SSIM
3 video sequences:
“Easy”, “Medium”,
“Complex”



Networks:



Based on data from:
J. Karlsson, and M. Riback. Initial field performance measurements of LTE, *Ericsson review*, 3, 2008.

- Quality-rate models: $Q(R) = \frac{R^{\beta}}{\alpha^{\beta} + R^{\beta}}$

Content	α	β
Easy	0.0555	0.8550
Medium	0.0724	0.8016
Complex	0.1015	0.7364

Network models: $p(R) = \alpha\, \mathcal{N}_{\mu_1,\sigma_1}(R) + (1-\alpha)\, \mathcal{N}_{\mu_2,\sigma_2}(R)$

Network	α	μ_1	σ_1	μ_2	σ_2
Network 1	0.584	0.996	0.564	2.554	1.165
Network 2	0.584	1.992	1.129	5.108	2.331

CONTEXT-AWARE ENCODING: EXAMPLE RESULTS

Optimal ladders for Network 1:

Content	N	Ladder bitrates [kbps]	Q_n	$\bar{Q}$	ξ [%]
Easy	2	138, 803	0.909	0.867	6.58
	3	100, 512, 1209	0.931	0.888	4.35
	4	100, 411, 866, 1645	0.946	0.897	3.34
	5	100, 349, 694, 1155, 2087	0.955	0.902	2.76
Medium	2	175, 854	0.881	0.830	7.98
	3	100, 518, 1219	0.906	0.854	5.31
	4	100, 416, 876, 1663	0.924	0.866	4.00
	5	100, 354, 701, 1165, 2104	0.936	0.873	3.25
Complex	2	234, 931	0.825	0.769	10.2
	3	145, 590, 1304	0.867	0.797	6.96
	4	102, 431, 898, 1704	0.888	0.812	5.22
	5	100, 363, 716, 1183, 2134	0.904	0.821	4.16

Optimal ladders for Network 2:

Content	N	Ladder bitrates [kbps]	Q_n	$\bar{Q}$	ξ [%]
Easy	2	232, 1457	0.940	0.906	5.14
	3	116, 811, 2124	0.955	0.924	3.27
	4	100, 589, 1421, 2803	0.964	0.932	2.40
	5	100, 486, 1107, 1974, 3577	0.971	0.937	1.92
Medium	2	293, 1549	0.920	0.878	6.23
	3	158, 893, 2216	0.939	0.899	4.04
	4	100, 601, 1438, 2828	0.949	0.909	2.97
	5	100, 495, 1123, 1995, 3615	0.958	0.915	2.35
Complex	2	391, 1685	0.887	0.833	7.98
	3	232, 1018, 2358	0.910	0.857	5.29
	4	156, 712, 1569, 3001	0.924	0.869	3.94
	5	114, 537, 1179, 2060, 3727	0.935	0.877	3.11

where:

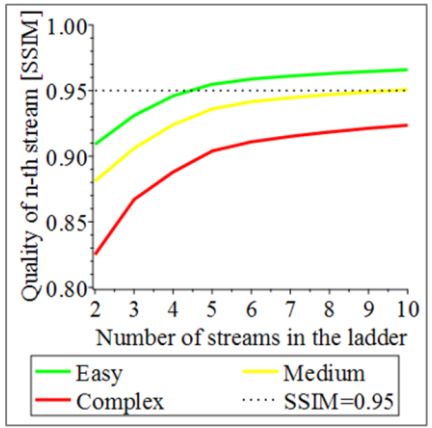
- Q_n = quality at top rendition [SSIM]
- $\bar{Q}$ = average quality [SSIM]
- ξ = gap to average quality achievable with infinite number of renditions [%]

- Key observation:
  - **optimal profiles designed for different sources and networks are different!**

CONTEXT-AWARE ENCODING: HOW MANY STREAMS ARE NEEDED?

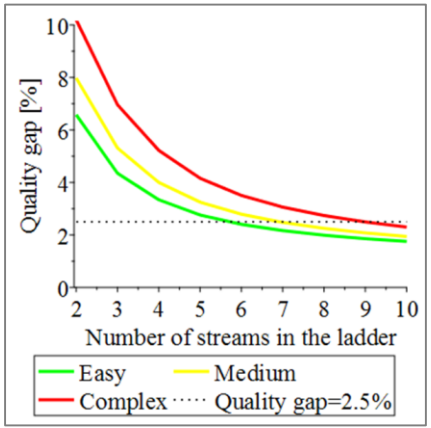
- There are two natural limits:

(1) Set limit for quality at top rendition:



This shows that "easy" content can be encoded with much fewer renditions!

(2) Set limit for quality gap:

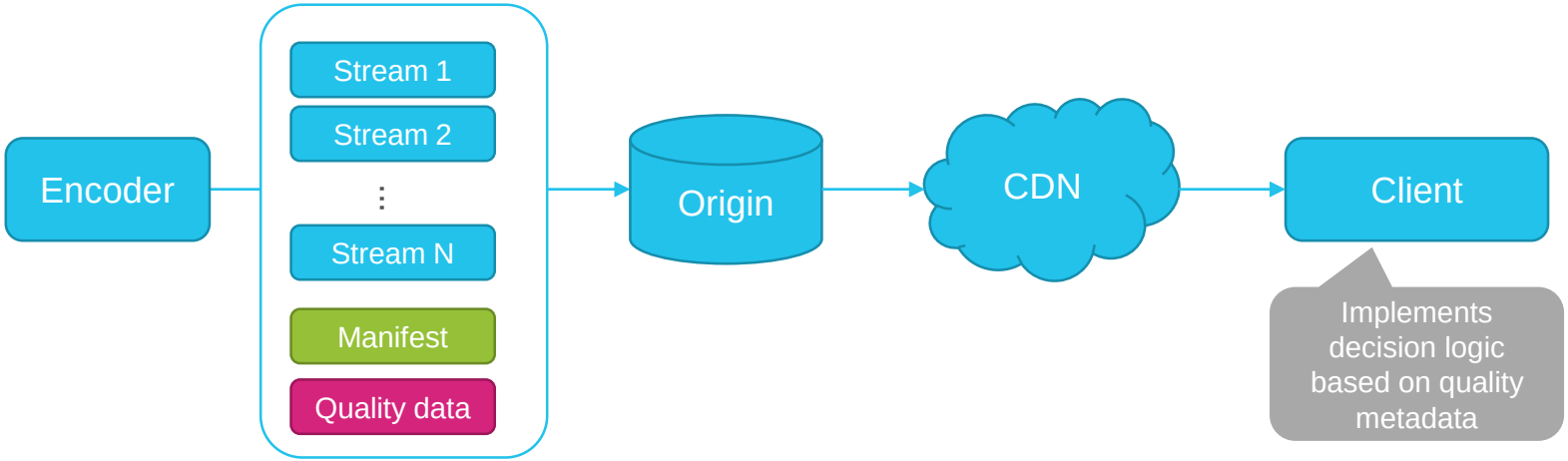


This provides effective bound on the number of renditions for "complex" content as well.

- This way, **the problem of design of optimal profiles for single codec case is now fully defined**:
  - we know how to choose rates & number of streams
  - best choices of resolutions follow by applying resolution-specific quality-rate functions

CAE APPROACHES: “QUALITY-BASED STREAMING”

- Architecture:



- Pros:
  - Allows per-segment adaptation
  - Allows clients to use advanced user- and context-aware adaptation strategies
- Cons:
  - Requires modifications of the standard
  - Quality-adaptive work in MPEG-DASH has not provided exact mechanism for enabling it

CAE TYPES: SUMMARY

CAE type	Example solutions	What it affects	Impact on standards
VBR encoding	x264/x265 “capped CRF”, Beamr CABR, Harmonic EyeQ, Elemental QVBR	Encoders Players need to be tested to operate reliably under VBR streams	Clarifications on the extent of VBR variation allowed may be useful.
Per-title encoding	Netflix per-title, Ateme CAE, Bitmovin’ per-title, Cambria, etc.	Encoders only Streams can be CBR or VBR	None
Context-aware encoding	Brightcove CAE, EpicLabs LightFlow, Mux “audience-aware encoding”	Encoders only Streams can be CBR or VBR	None
Per-scene encoding	Netflix per-scene encoding	Encoders, players	Needs seamless multi-period option (ability to switch to new manifest on a per-scene basis)
Quality-based streaming	MediaMelon QBR, Bitmovin’ per-scene adaptation	Encoders, players	Needs exact means of signaling of quality annotations and definition of anticipated client behavior (both in cases of quality-aware and legacy clients).

DISCUSSION TOPICS: CAE AND 5G

Main Questions:

- **CAE / ABR ladder design and 5G:**
  - is there an overlap?
  - if networks are improving, do we still need ABR?

- **5G network characteristics and their impact on streaming:**
  - are there any significant differences in shape of network throughput CDF in 5G vs older networks?
  - is there any way clients can be advised by the core network about current load and hence shape of network bandwidth PDFs and other relevant statistics?

DISCUSSION TOPICS: CAE AND STANDARDS

Two CAE architectures likely need standards support:

- **Per-scene encoding:**
  - this requires clients to be able to adapt to a new encoding manifest provided on a per-scene basis
  - what is needed is basically a “seamless multi-period” option in DASH
    - could be constrained to: same codec, same number of streams, but bitrates will definitely be different

- **Quality-driven streaming**
  - this needs exact means of signaling of quality annotations
    - MPEG-B “carriage of timed metadata” spec is a good start, but its use for the purpose is not defined anywhere
  - what also needs to be understood and enabled is **backwards-compatible regime of operation**
    - If new clients know nothing about quality metadata, they must be able to deliver same content as reliably as new players, but perhaps less efficiently.

THANK YOU!

YURIY REZNIK | YREZNIK@BRIGHTCOVE.COM