MediaKind

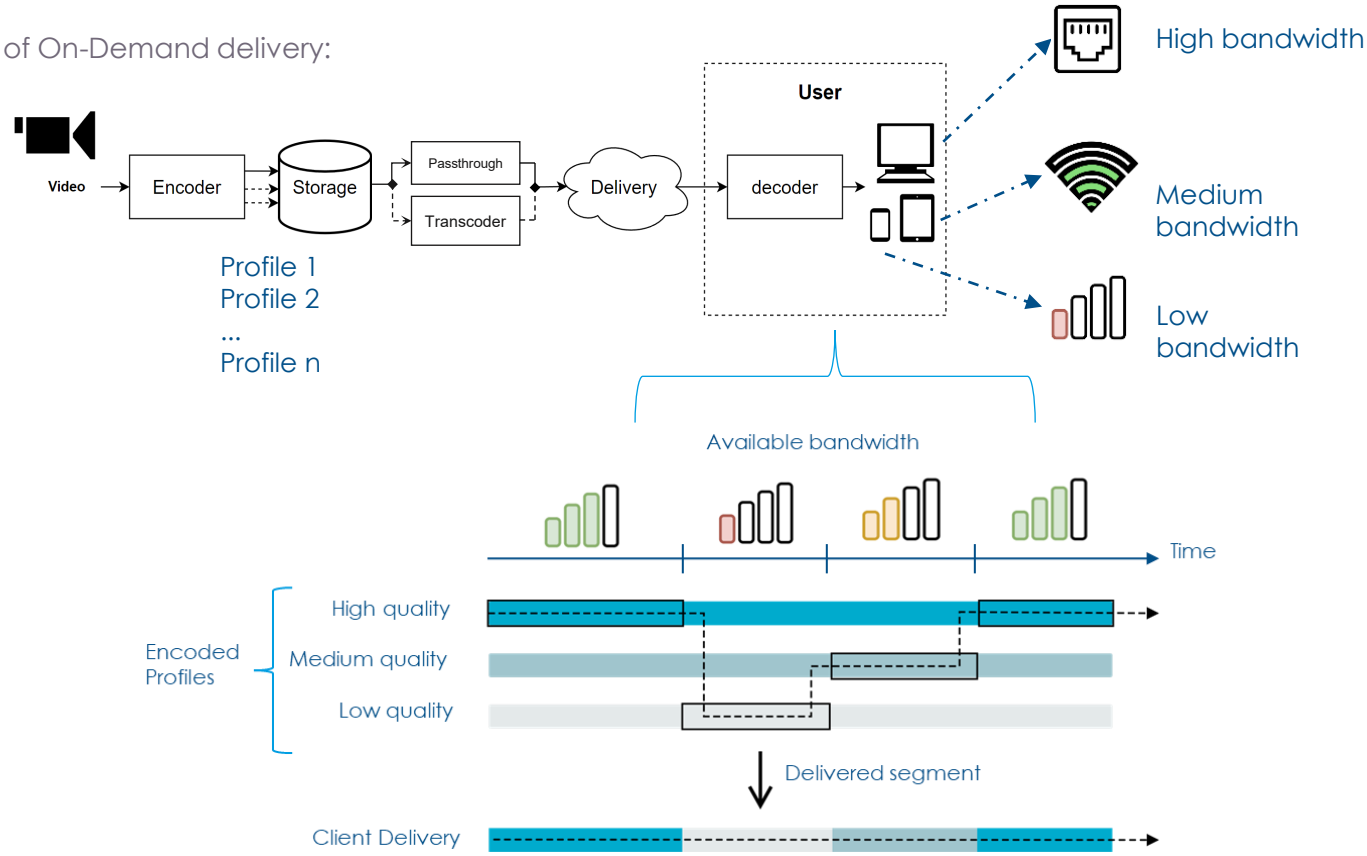
Optimizing ABR Streaming Costs with Joint Multi-Profile Coding

Paul O’Donovan – Senior Product Manager

Paul.ODonovan@MediaKind.com

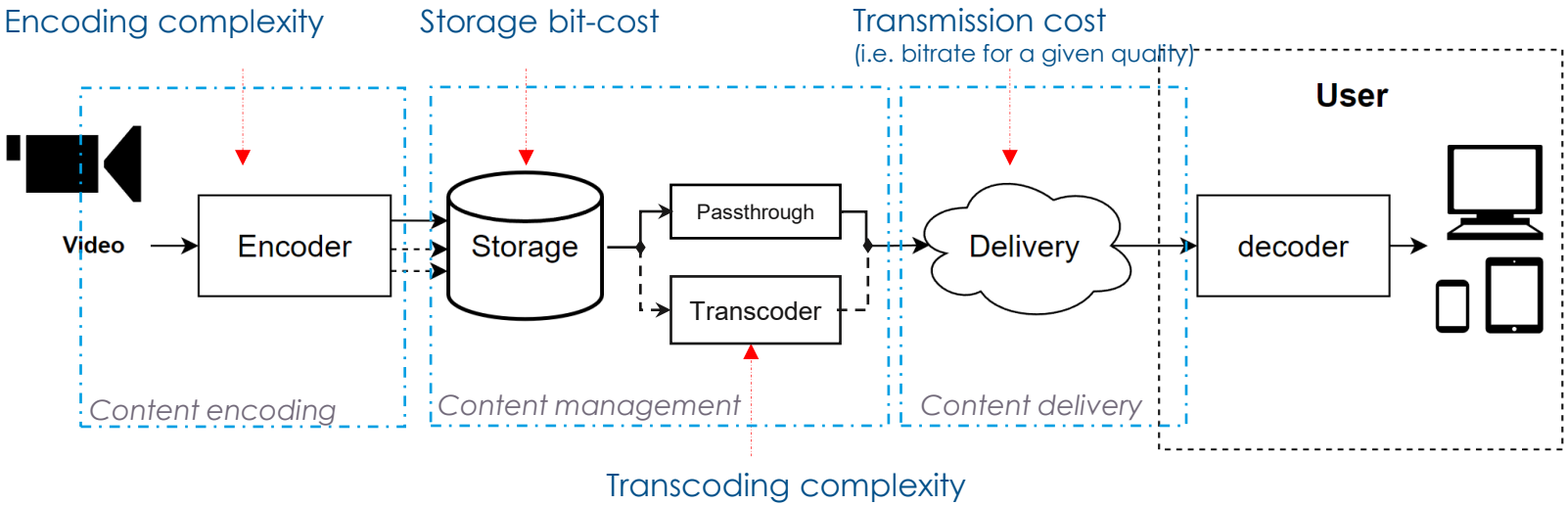
Context: Adaptive Bitrate Streaming (ABR)

Example of On-Demand delivery:



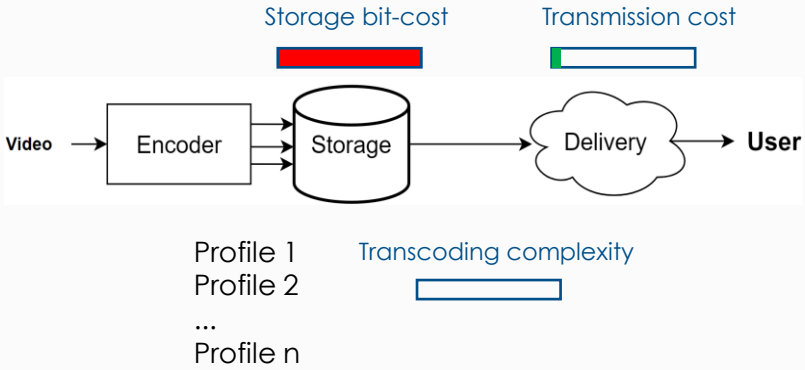
ABR delivery system and associated costs

Example of On-Demand delivery:

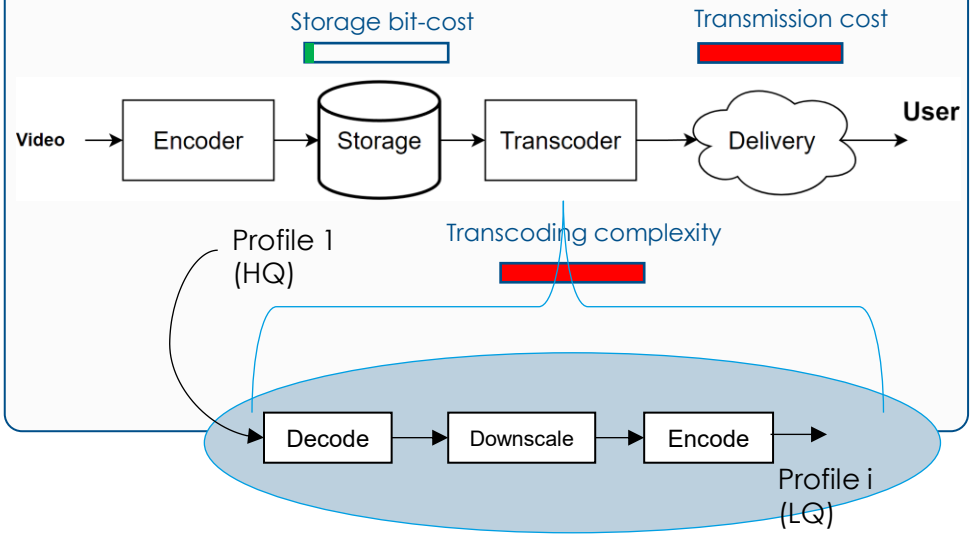


ABR system optimization: two common extrema for delivery

Simulcast (SC) delivery

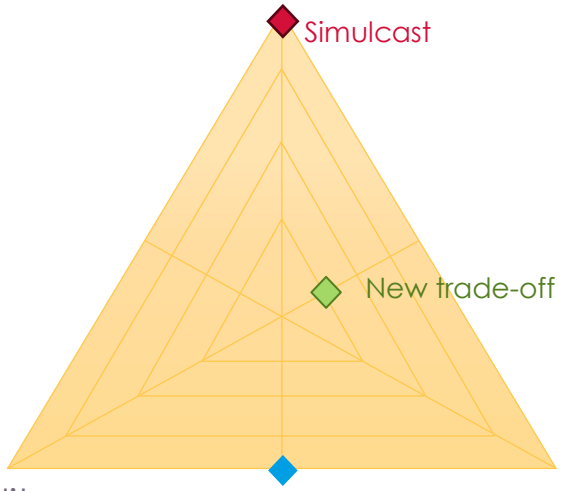


Full Transcoding (FT) delivery



Investigating better trade-off ?

Storage (or delivery) **cost**
of all profiles



Transmission cost
i.e. bitrate for a given quality
of the requested profile

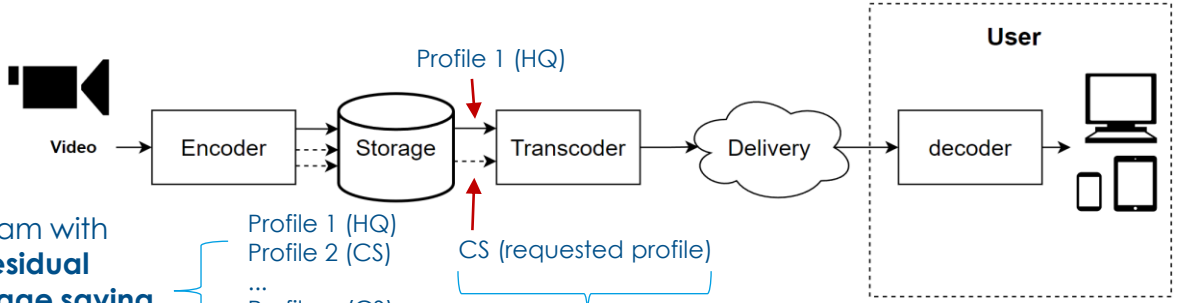
Transcoding complexity
for generating the
requested profile

with the operational constraints that:

- the **delivered stream** remains **compliant with standard decoder** available at the client
- the **highest quality profile** has the **same transmission cost and quality** than for Simulcast

State-of-the Art: Guided Transcoding

- Introduce the concept of Guided Transcoding by means of Control Stream (CS)



CS = LQ stream with **pruning of residual data for storage saving** (i.e. transformed coefficients)

Residual data are regenerated on client-request using the decoded HQ profile as input, and prediction syntaxes from the CS **saving heavy coding mode/prediction search**

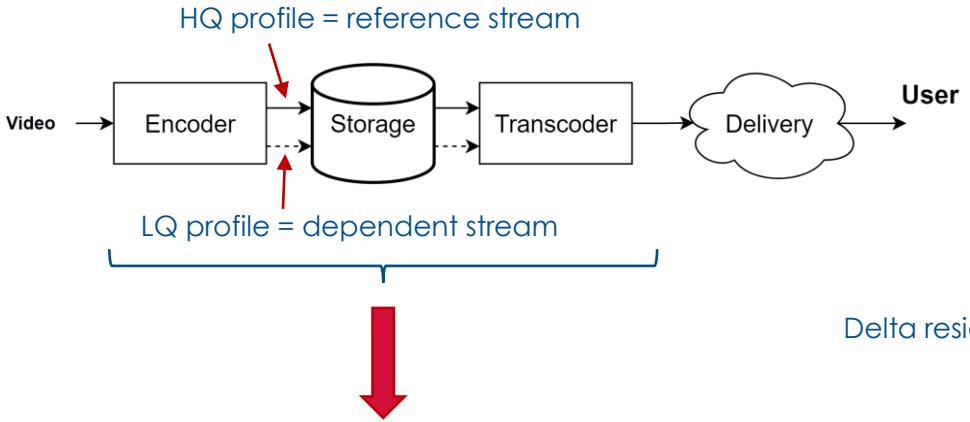
Storage bit-cost < Simulcast

Transcoding complexity < Full Transcoding

Transmission cost = Full Transcoding

State-of-the Art: Guided Transcoding with Deflation and Inflation

- Hollmann et al., 2018
- Introduce the principle of Predictive Residual Coding (PRC) across profiles for storage saving



Dependent stream → **Original** residual

\-

Reference stream → **Predicted** residual

=

Delta residual

Delta residual < Original residual → **Storage savings for LQ streams**

= "Deflation" process

State-of-the Art: Guided Transcoding with Deflation and Inflation

Advantages

- Lower transcoding complexity vs Full transcoding

- Lower storage requirements vs Simulcast

- Transmission cost equal to Simulcast

Disadvantage

- Complex residual predictor generation
  - Requires 2 full decoding loops for transcoding

Contribution: further **reduce transcoding complexity by using 1 partial decoding loop for predictive residual coding** at the cost of lower storage saving

Contribution: coding efficiency optimizations

Introduce two complementary optimizations for coding efficiency improvement of any method based on Predictive Residual Coding

1. Conditional Delta Residual (CDR) coding and signaling
   - Motivation: discard residual predictor not well correlated with the residual samples to predict
   - For every coding unit (CU), code the delta residual only if it lowers the bit-cost, else code the original residual
   - Add a 1-bit CU flag for signaling the decision

2. Rate-Distortion Optimization based on Delta Residuals (RDODR)
   - Motivation: favor prediction and coding modes that will minimize the delta residual to code for the dependent streams
   - Update the RDO process used for coding mode/prediction search by using delta residual bit-cost for the rate estimations

Experimental framework and test conditions

- Proposals and State of the Art (SOTA) methods compared in the context of VVC codec
  - Implementations done on top-of VTM-19.0 leveraging on VVC Multi-layer coding structure
  - Including GTDI method renamed as PRC-Full-PTQ in subsequent tables

- Performance assessment vs. Simulcast (SC) and Full Transcoding (FT) in terms of:
  - Storage bit-cost, transmission efficiency and transcoding complexity
  - Using similar test condition and test sequences than the MPEG CfE on NDVC

Sequence ID	Sequence name	Frame count	Frame rate	Bit depth	Resolution*	Subsampled resolutions
UHD1	CatRobot1	600	60fps	10	3840x2160	1440p,1080p,720p,540p,360p
UHD2	FoodMarket3	720	60fps	10	3840x2160	1440p,1080p,720p,540p,360p
UHD3	BuildingHall1	500	50fps	10	3840x2160	1440p,1080p,720p,540p,360p
HD1	Cactus	500	50fps	8	1920x1080	720p, 540p, 360p
HD2	BQTerrace	600	60fps	8	1920x1080	720p, 540p, 360p
HD3	BasketballDrive	500	50fps	8	1920x1080	720p, 540p, 360p

- For two video delivery scenarios: multi-bitrate and multi-resolution

Experimental results: multi-bitrate scenario

		Storage Bitcost vs SC				Transmission efficiency vs SC		Transcoding complexity vs FT	Transcoding complexity vs PRC-Full-PTQ
	Streams:	All		Dependent		Dependent		Dependent	Dependent
Approach	Variant	QP_0 22	QP_0 27	QP_0 22	QP_0 27	QP_0 22	QP_0 27	QP_0 22, 27	QP_0 22, 27
SOTA									
Full Transcoding* [10]	NA	-60.5%	-66.1%	-100%	-100%	14.2%	17.5%	0%	2079.8%
PRC-Full-PTQ	Base* [11]	-25.3%	-25.5%	-41.4%	-38.5%	0%	0%	-95.2%	0%
Proposals									
	CDR	-24.8%	-24.4%	-40.5%	-36.8%	0%	0%	-95.2%	0%
	CDR+RDODR	-29.0%	-30.3%	-47.3%	-45.7%	0.4%	-0.9%	-95.2%	0%
PRC-Part-TQ	Base	-6.4%	-6.9%	-10.5%	-10.5%	0%	0%	-98.5%	-67.5%
	CDR	-7.9%	-7.8%	-13.0%	-11.8%	0%	0%	-98.5%	-67.5%
	CDR+RDODR	-10.7%	-10.8%	-17.6%	-16.4%	1.8%	0.8%	-98.5%	-67.5%

Guided Transcoding with Deflation and Inflation:

- About -40% storage saving for dependent streams (-25% overall) vs. Simulcast
- Same transmission cost than Simulcast; no bitrate overhead for the same quality
- -95% run-time reduction vs. Full Transcoding
- Proposed R-D optimizations can boost storage saving by -8% for negligible impact on transmission cost

Conclusion

- Significant cost savings are possible using new Guided Transcoding techniques
  - vs. Simulcast and Full Transcoding
  - Works with any client
- Approaches are compatible with:
  - Content- or Audience- aware dynamic bitrate ladder
  - CDN optimization
- Future use cases include VOD archive and cDVR services.

MediaKind.com