CODIFICAÇÃO POR TÍTULOS: OS SEGREDOS POR TRAZ DO OTT DA NETFLIX

Neste artigo, vou procurar mostrar o histórico das tecnologias de codificação por título, como forma de dar subsídios de quais recursos devem ser procurados ao escolher uma tecnologia e/ou provedor de serviços.

Por: Tom Jones Moreira

Primeiro uma explicação: O que é codificação por titulo? Uma codificação regular simplesmente segue as regras que o usuário define em suas configurações de *stream*, codec e *muxing*. Uma codificação por título em comparação, não apenas usa a configuração fornecida, mas avalia o próprio recurso. Com base nesses dados, o algoritmo ajusta os parâmetros de largura, altura e taxa de bits, otimizando assim a sua escala de taxa de bits para aumentar a qualidade e, ao mesmo tempo, reduzir a largura de banda necessária para entregá-la.

Então o objetivo da codificação por título é otimizar a escala de taxa de bits para cada codificação, e então tentar encaixar o resultado na faixa relativamente estreita que o sistema perceptivo humano pode ver. Portanto, existe uma balança: Se pendermos para além do sistema visual humano se perderam muitos bits em codificações. Porém se formos abaixo disso, criaremos muitos artefatos que serão perceptíveis aos usuários. Em última análise, o que a codificação por título nos permite fazer é tentar reduzir o custo e melhorar a qualidade que podemos colocar em cada codificação.

O **perfil por título** é capaz de alcançar uma qualidade muito melhor usando bits rates mais baixos na faixa de percepção BRLBTC humana (aproximadamente entre 35dB e 45dB). Além disso, remove as tuplas de resolução/taxa de bits acima de 45dB, o que não leva mais a uma melhoria da qualidade visual. Dito isso, podemos seguir.

O que começou como um ajuste de taxa de dados unidimensional que refletia a realidade simples de que todos os vídeos eram codificados de maneira diferente, agora é uma análise complexa que incorpora taxa de quadros, resolução, gama de cores e faixas dinâmicas, bem como redes de distribuição e dados relacionados ao dispositivo. Ao longo do caminho, as métricas de qualidade de vídeo (VQM) também avançaram para fornecer os dados relacionados à qualidade que alimenta os algoritmos por título.

A ideia do texto é procurar mostrar o histórico das tecnologias de codificação por título, como forma de dar subsídios de quais recursos devem ser procurados ao escolher uma tecnologia e/ou provedor de serviços. Embora a **Netflix** seja geralmente creditada com a invenção da codificação por título, desde que publicou seu artigo em dezembro de 2015, intitulado "Otimização de codificação por título" (1). A verdade é que várias tecnologias ja existiam antes disso, como por exemplo: *Content Adaptive Bitrate* (CABR) e *Constant Rate Factor* (CRF) da Beamr (2), que hoje estão presentes em codecs como H264, H265 e VP9.

O gráfico abaixo compara o chamado *Convex Hull* (basicamente um envelope das regiões onde certas combinações de resolução/taxa de bits funcionam melhor) de uma escala de taxa de bits padrão (não dinâmica) e por título.

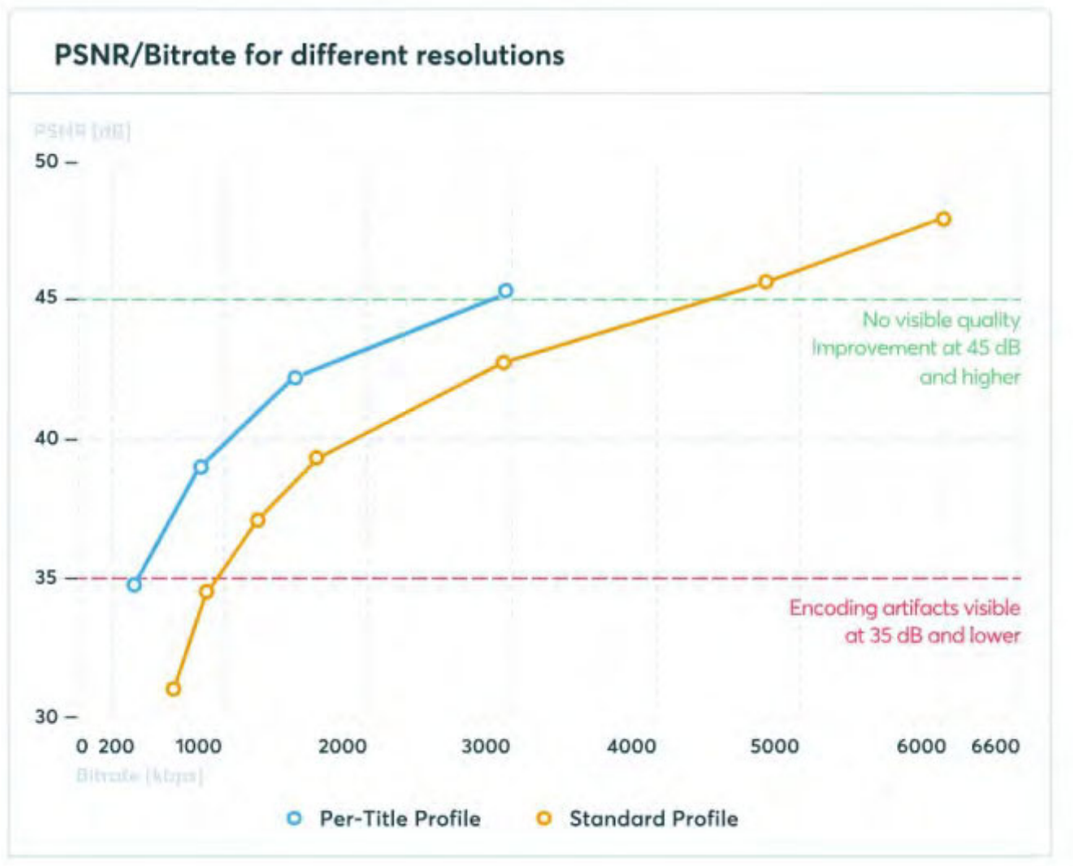

Fig.1 - Comparação entre título e perfil padrão em termos de qualidade e taxa de bits
Fonte: https://bitmovin.com/docs/encoding/faqs/what-is-per-title-encoding

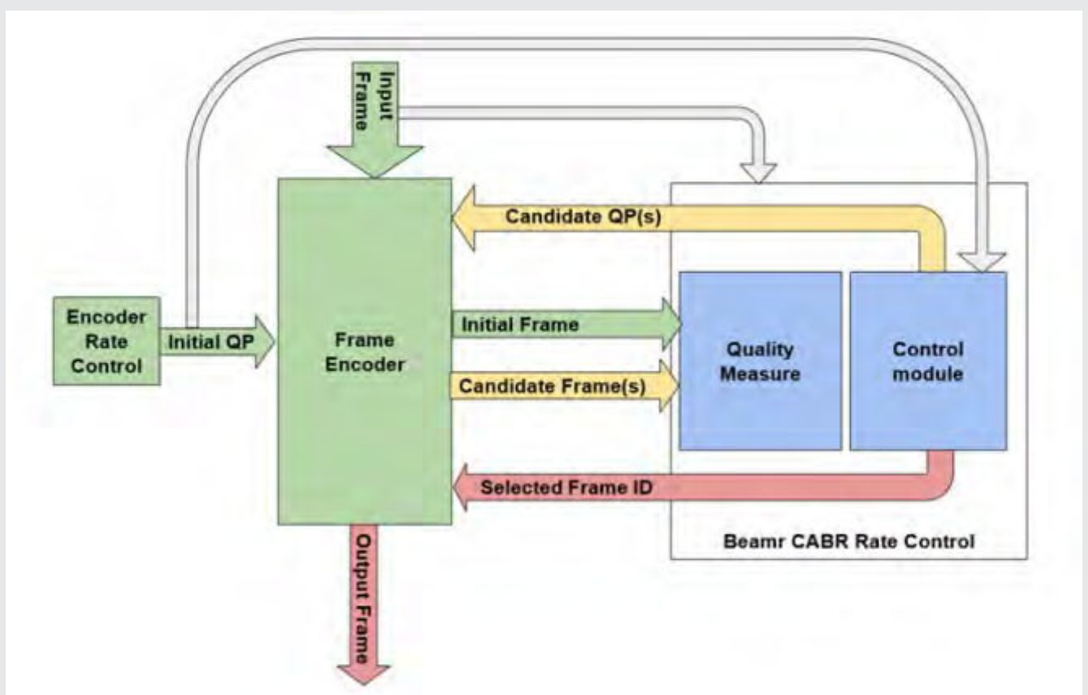

Fig.2 – Diagrama do CABR do Beamr (Fonte: www.beamr.com)

Em geral, as tecnologias de otimização são a única maneira prática de ajustar a taxa de dados do arquivo durante eventos ao vivo. Por esse motivo, as tecnologias de codificação ao vivo como a *Constant Rate Factor* da AWS Elemental, ou a *Quality-Defined Variable Bitrate* (QVBR), ou ainda a EyeQ da Harmonic sãos chamadas de tecnologias de otimização. Porém nem tudo são flores, sabemos hoje que todas as tecnologias de otimização têm uma limitação muito séria: elas não podem alterar nenhum aspecto do arquivo ou da escala de codificação além da taxa de bits. Sendo bem objetivo, observe que sempre que avaliamos uma tecnologia de codificação por título, normalmente a comparamos com uma escala de codificação fixa.

Conforme demonstrado na Tabela 1, com a escala fixa à esquerda e duas tecnologias por título à direita. Vemos que a tabela A é uma tecnologia de otimização; isso é observado uma vez que ao alimentarmos a tabela com o número de variações de resoluções da tabela de codificação original, vemos como resultado direto que o algoritmo reajusta a taxa de dados - e apenas a taxa de dados - desses diversos “degraus” da tabela original.

Em contraste a tudo isso, ao alimentarmos a tabela da tecnologia B, com o mesmo arquivo original, vemos que ela decide quantos degraus a escala de codificação precisa, bem como qual sua taxa de dados e resolução. Na figura, podemos ver que a tecnologia B, não apenas reduz o número de degraus (e os custos de codificação), mas também aumenta as resoluções desses degraus e a qualidade de vídeo associada, a isso chamamos: *Multimethod Assessment Fusion* (VMAF). Como as tecnologias de otimização só podem ajustar a taxa de dados, não o número de degraus ou sua resolução, elas normalmente não funcionam tão bem, quanto outras tecnologias por título, que podem ajustar todas as três variáveis. Ainda assim, as tecnologias de otimização eram tudo o que existia até a Netflix estrear sua tecnologia de codificação por título, em dezembro de 2015.

TABELA ORIGINAL

	Width	Height	BitRate	PSNR	SSIM	VMAF
1080p_CVBR	1920	1080	4.433	42.64	0.958	94.79
720p_CVBR	1280	720	2.677	40.95	0.954	89.89
540p_CVBR	960	540	1.895	39.83	0.950	85.72
480p_CVBR	854	480	1.350	39.20	0.947	82.80
360P_CVBR	640	360	897	37.97	0.939	74.18
270p_CVBR	480	270	491	35.95	0.924	54.77
180p_CVBR	320	180	232	32.95	0.898	20.49

TECNOLOGIA A

	Width	Height	BitRate		PSNR	SSIM	VMAF
1080p_arqxyz1.mp4	1920	1080	859.9	2.07	49.25	0.995	96.79
1080p_arqxyz2.mp4	1280	720	415.7	1.94	45.09	0.995	95.94
720p_arqxyx1.mp4	960	540	214.8		35.05	0.988	90.22

TECNOLOGIA B

	Width	Height	BitRate		VMAF
1080p_CVBR	1920	1080	375	1.93	95.68
720p_CVBR	1280	720	194	1.65	88.33
540p_CVBR	960	540	118	1.20	81.15
480p_CVBR	960	540	99	1.57	77.75
360P_CVBR	854	480	63	1.51	64.43
270p_CVBR	640	360	42	1.90	41.38
180p_CVBR	480	270	22		4.24

Tabela 1: Tecnologias

Ela usa uma técnica de codificação de força bruta, que codifica cada arquivo de origem em centenas de combinações de resolução e taxa de dados para encontrar o que a Netflix chama no artigo de *"Converx Hull",* que limita de forma mais eficiente todos os pontos de dados, conforme demonstrado na Figura 3.

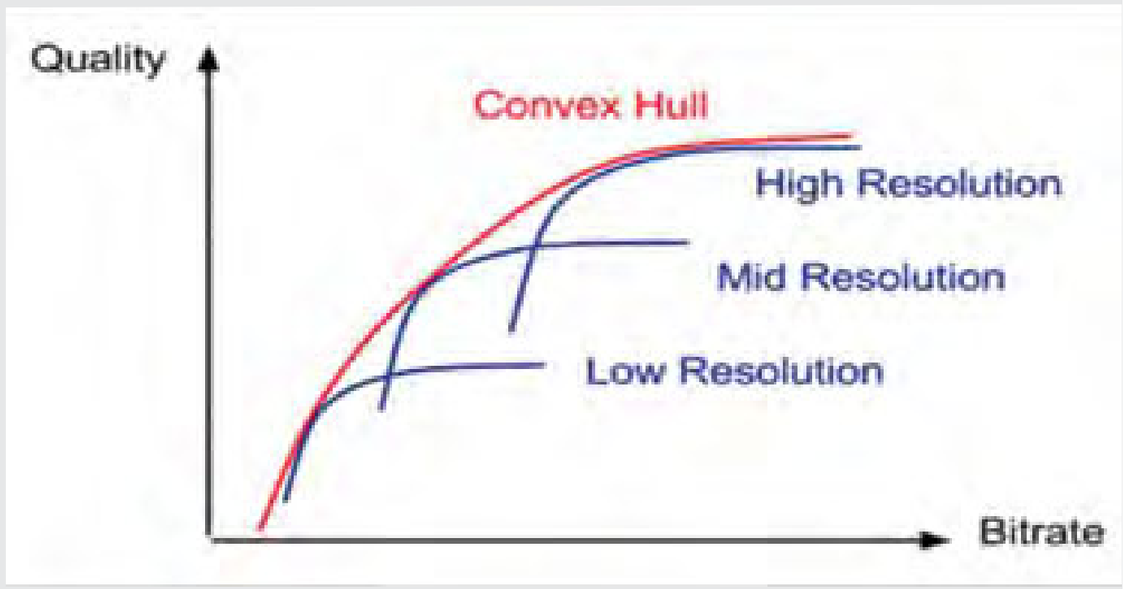

Figura 3 –Convex Hull (Fonte: netflixblog.com)

Curiosamente, a métrica que originalmente impulsionou o mecanismo de decisão da Netflix foi a relação sinal-ruído de pico (PSNR), que é uma métrica de imagem estática que não incorpora o conceito de movimento. A Netflix substituiu o PSNR pelo VMAF em junho de 2016. Resumidamente, o VMAF combina quatro métricas de qualidade, incluindo uma métrica de movimento simples.

Quando lançado, a plataforma postou dados mostrando que o VMAF tinha uma correlação mais alta com avaliações subjetivas do que o PSNR. No início de 2016, ficou claro que muitas outras organizações vinham trabalhando na implementação por título há algum tempo, o próprio YouTube apresentou um artigo que detalhou sua abordagem para o problema de forma bem diferente. Enquanto a Netflix codifica comparativamente poucos vídeos, mas a maioria é assistida por milhões de clientes pagantes, o que justifica seu caro esquema de codificação que oferece a melhor qualidade absoluta com a menor taxa de bits possível.

O YouTube estava recebendo aproximadamente mais de 300 horas de vídeo por minuto, com alguns vídeos sendo assistidos por milhões de telespectadores, mas a maioria assistida por números bem menores. Isso exigia uma implementação por título muito mais rápida e econômica. Curiosamente, a técnica do YouTube combinou Inteligencia Artificial (IA) com dados de complexidade de arquivos fornecidos por uma único arquivo de origem codificado em CRF 240p. Em 2018, a Netflix deu mais um passo afrente e estreou seu *Dynamic Optimizer* baseado em cena. Conforme mostrado na Figura 3, em vez de dividir o vídeo em GOPs ou segmentos arbitrários de 2 ou 3 segundos, a otimização dinâmica divide o vídeo em cenas e codifica cada cena separadamente. Enquanto isso cria um GOP dinâmico e comprimentos de segmento, a comutação de fluxo de taxa de bits adaptável (ou *Adaptive Bit Rate* - ABR) continua a funcionar efetivamente porque todos os degraus de escala compartilham o mesmo GOP e comprimentos de segmento.

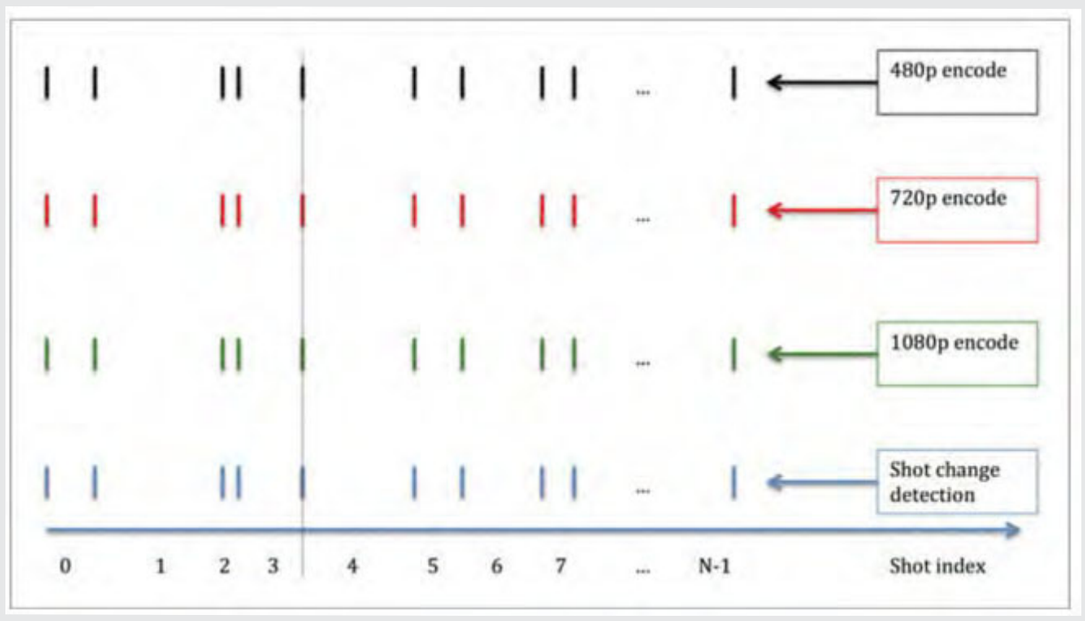

Figura 4- Otimização Dinâmica baseada em cenas (Fonte: Netflixblog)

Intuitivamente, a codificação baseada em cena faz muito sentido. A personalização dos parâmetros de codificação para uma cena individual deve ser mais eficiente do que tentar encontrar a configuração de codificação ideal para um segmento que contém duas ou mais cenas que podem incluir um conteúdo totalmente diferente. Além disso, alterar os parâmetros de codificação em uma mudança de cena obviamente seria menos perceptível do que alterar os parâmetros em uma cena.

Por fim, os números de eficiência computados pelas Netflix demonstraram; conforme medido pelo VMAF, que a otimização dinâmica permitiu a redução das taxa de bits de H264, VP9 e H265 em 28,04%, 37,61% e 33,51%, respectivamente, mantendo a mesma qualidade. Embora o título baseado em tomadas seja atraente, ele cria problemas significativos no lado do *player* ou decodificador, especialmente para aplicativos que envolvem a inserção de publicidade. No mínimo, o usuário precisará de players/ aplicativos personalizados em praticamente todas as plataformas e, mesmo assim, pode ser muito complicado inserir anúncios quando mais precisar. Definitivamente, a melhor estratégia aqui é verificar qual o *status* dos players/decoders antes de começar a mexer no lado da codificação.

O próximo avanço veio de uma direção completamente diferente e responde afirmativamente à seguinte pergunta: **"Você criaria sua escala de codificação de forma diferente se soubesse quais dispositivos estão reproduzindo seu conteúdo e em quais velocidades de conexão?"**. Técnicas de codificação por título que incorporaram dados de reprodução surgiram de três empresas diferentes, quase que ao mesmo tempo: Brightcove, Mux e Epic Labs, agora de propriedade da Haivision. A melhor descrição é fornecida em um *white paper* intitulado "Otimizando a entrega de vídeo em várias telas em grande escala"(3), de autoria *Yuriy Reznik* e outros três colegas da *Brightcove.*

O documento descreve a tecnologia *Context-Aware Encoding (CAE) da Brightcove,* que analisa o conteúdo "e as estimativas das probabilidades de carregamento do fluxo em cada taxa para cada cliente". E continua, "no cálculo da expressão de custo de otimização final, o gerador CAE agrega estimativas obtidas para cada tipo de cliente de acordo com a distribuição de uso, também fornecida pelo módulo analítico. Em outras palavras, a geração de perfil CAE é realmente um processo de otimização de ponta a ponta para entregar em vários dispositivos e telas".

O artigo analisa os três padrões de uso mostrados à esquerda na Figura 4 e apresenta a escala de codificação exclusiva criada para cada padrão de uso do mesmo conteúdo. O primeiro padrão de uso é centrado em dispositivos móveis, o segundo é de uso mais geral e o terceiro é centrado em IPTV, com 100% de toda a distribuição para TVs com uma largura de banda média de cerca de 36 Mbps. Embora a diferença entre as duas primeiras escalas de codificação seja muito sutil, a terceira escala se destaca por ter o menor número de degraus, a menor taxa de bitrate total, além de ter o degrau superior de mais alta qualidade. Isso reduz os custos de codificação e armazenamento e melhora a qualidade da experiência (QoE).

Device type	Usage [%]	Average bandwidth [Mbps]
PC	0.004	7.5654
Mobile	94.321	3.2916
Tablet	5.514	3.8922
TV	0.161	5.4374
All devices	**100**	**3.3283**

TABLE 2: USAGE AND AVERAGE BANDWIDTH STATISTICS FOR OPERATOR 1.

Device type	Usage [%]	Average bandwidth [Mbps]
PC	63.49	14.720
Mobile	6.186	10.609
Tablet	9.165	12.055
TV	21.15	24.986
All devices	100	16.393

TABLE 3: USAGE AND AVERAGE BANDWIDTH STATISTICS FOR OPERATOR 2.

Device type	Usage [%]	Average bandwidth [Mbps]
PC	0.0	N/A
Mobile	0.0	N/A
Tablet	0.0	N/A
TV	100	35.7736
All devices	100	35.7736

TABLE 4: USAGE AND AVERAGE BANDWIDTH STATISTICS FOR OPERATOR 3.

Tabela 2: Tabelas de comparação
(Fonte: https://www.researchgate.net/publication/333040039)

Rendition	Profile	Resolution	Framerate	Bitrate	SSIM
1	Baseline	320x180	30	125	0.93369
2	Baseline	480x270	30	223.08	0.93793
3	Main	640x360	30	398.11	0.94636
4	Main	960x540	30	774.78	0.94953
5	Main	1280x720	30	1549.5	0.95637
6	High	1600x900	30	2765.3	0.96105
7	High	1920x1080	30	4935.1	0.96576
Storage				10771	

TABLE 9: CAE-GENERATED ENCODING LADDER FOR OPERATOR 1.

Rendition	Profile	Resolution	Framerate	Bitrate	SSIM
1	Baseline	320x180	30	125	0.93338
2	Baseline	480x270	30	239.71	0.94122
3	Main	640x360	30	469.54	0.95202
4	Main	1024x576	30	939.08	0.95221
5	Main	1280x720	30	1568.8	0.95658
6	High	1600x900	30	2765.3	0.96105
7	High	1920x1080	30	4935.1	0.96576
Storage				11026	

TABLE 10: CAE-GENERATED ENCODING LADDER FOR OPERATOR 2.

Rendition	Profile	Resolution	Framerate	Bitrate	SSIM
1	Baseline	320x180	30	125	0.93447
2	Baseline	512x288	30	307.42	0.94855
3	Main	960x540	30	803.59	0.95050
4	Main	1280x720	30	1727.8	0.95864
5	High	1920x1080	30	5050.7	0.96599
Storage				8014.6	

TABLE 11: CAE-GENERATED ENCODING LADDER FOR OPERATOR 3.

Tabela 3: Tabelas de comparação
(Fonte: https://www.researchgate.net/publication/333040039)

Outra variável abordada no artigo da Brightcove são as implementações de codecs múltiplos. Aqui, o artigo afirma: “Uma das características do gerador de perfil CAE é a capacidade de gerar perfis ABR para a pluralidade de codecs existentes. Nesse caso, o gerador também usa informações sobre o suporte de tais codecs por diferentes categorias de dispositivos receptores. Essas informações são fornecidas como parte das estatísticas de uso do operador e largura de banda, fornecidas pelo mecanismo de análise.”

O uso da geração de perfis multi-codec leva a economias adicionais no número total de renderizações e ganhos de qualidade alcançáveis por clientes que podem alternar entre os codecs (por exemplo, H264 e HEVC). Incorporar vários codecs em uma única escala também faz muito sentido, a produção de uma única escada híbrida economiza custos de codificação e armazenamento, tornando-a a melhor opção para a maioria dos produtores que oferecem ambos os codecs, tornando a capacidade de criar, escalas de codificação de codec híbrida, um recurso valioso para tecnologias por título.

Embora as considerações de faixa dinâmica sejam comparativamente novas, os produtores têm reduzido as taxas de quadros dos degraus mais baixos de seus níveis de codificação já faz algum tempo. No entanto, nenhuma das tecnologias por título discutidas até agora, destacou a capacidade de ajustar automaticamente a taxa de quadros dos degraus da escala. Esse problema é agravado por vídeos de 50/60 fps, onde duas ou três opções de taxa de quadros podem ser necessárias para representar uma progressão suave da qualidade das taxas de bits mais altas, para as mais baixas.

Outro problema semelhante está relacionado à faixa dinâmica. Embora o HDR seja claramente preferível nos degraus superiores da escala de codificação, ele pode ser incompatível com os degraus inferiores. Dessa forma é importante que as técnicas de codificação por título (que tratam de conteúdo premium) com altas taxas de quadros e HDR, também devem abordar como escalar a taxa de quadros e a faixa dinâmica, junto com todos os outros parâmetros discutidos anteriormente, em uma única escala de codificação. E esse é um desafio enorme para os fabricantes de encoders, e podemos encontrar uma primeira tentativa de se fazer isso, ao olharmos para um paper da francesa ATEME, que aborda a taxa de quadros e a adaptação de faixa dinâmica, intitulado: "Codificação de Conteúdo voltado para a próxima geração de UHD, HDR, WCG, HFR" (4), de autoria de Thomas Guionnet e outros dois colegas.

O artigo explora o impacto da adaptação da taxa de quadros primeiro, usando o Índice de Qualidade ATEME (do inglês: ATEME *Quality Index-AQI*) para traçar as curvas mostradas na Figura 5, que rastreia a qualidade dos fluxos que variam em resoluções de 960x540 a 4K, e com taxas de quadros que variam de 25 a 100 fps, além de taxas de dados que variam cerca de 1 Mbps a 80 Mbps.

No artigo, os pesquisadores detalham que as métricas da ATEME incorporam resolução, taxa de quadros, faixa dinâmica e gama de cores; e essas técnicas são independentes dos codecs; pois usam a otimização de treliça para produzir os degraus ideais para a escala de codificação.

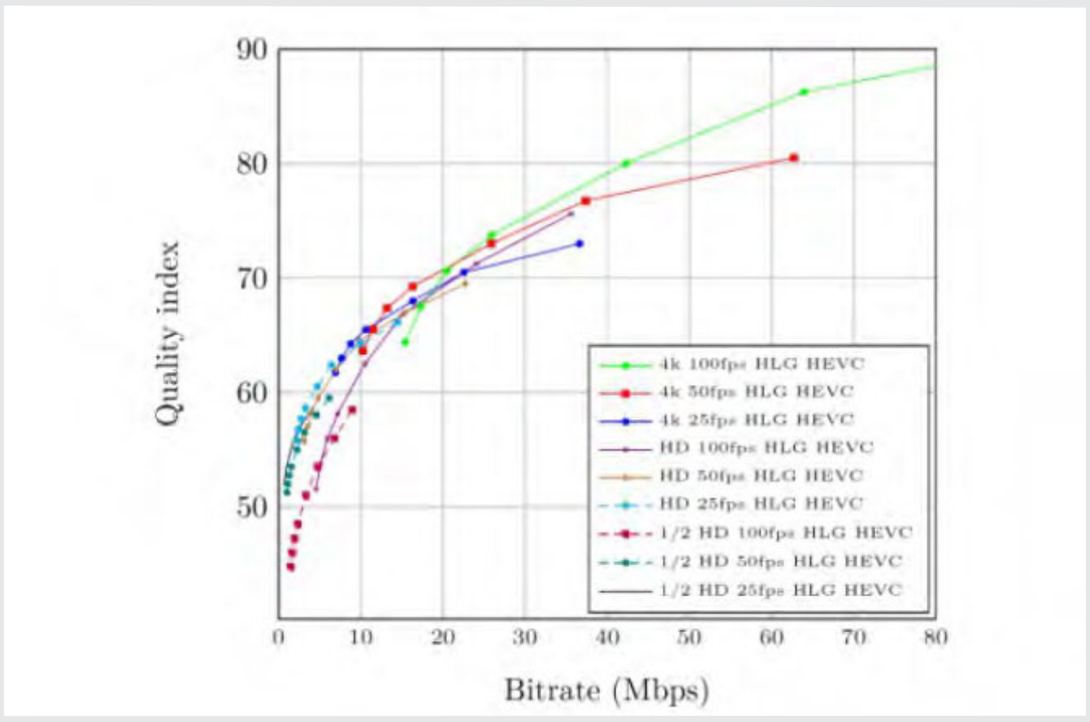

Figura 5 - Conjunto de curvas de qualidade de taxa para a variação de resoluções de 960x540 a 4K e framerate de 25 a 100 fps. (Fonte: Ateme)

Codec	Resolution	Fps	Dynamic range	Bitrate (kbps)
HEVC	7680x4320	50	HDR	17606
HEVC	3840x2160	50	HDR	6924
HEVC	2560x1440	50	HDR	3095
HEVC	1920x1080	50	HDR	1755
HEVC	1280x720	50	HDR	1054
HEVC	960x540	50	HDR	642
HEVC	640x360	25	HDR	383
HEVC	480x270	25	HDR	224

Tabela 4 – Conteúdo adaptado de conjunto de perfis para sequência Polynésie (Fonte: Ateme)

Por exemplo, o artigo sita que se trabalharmos com uma entrada de 8K a 50 fps HDR Hybrid *Log-Gamma* (HLG), no caso um vídeo da Polinésia, o sistema ira produzir uma escala de codificação, com resultados mostrados na Tabela 4, que delineia pontos de troca para taxa de dados e faixa dinâmica.

Os autores explicam: “Como ilustração de sua flexibilidade, o framework foi aplicado em duas sequências... O codec foi corrigido para HEVC e uma máxima resolução obrigatória em conjunto com uma taxa de bits mínima foram fornecidas como restrições. Dessa forma a resolução, taxa de quadros e a faixa dinâmica foram deixadas para que o framework as configura-se automaticamente...A escala de perfis parece normal, exceto para a resolução que abrange uma faixa muito ampla. A resolução vai diminuindo suavemente junto com a taxa de bits, enquanto a taxa de quadros é reduzida apenas para os perfis mais baixos.”

Os autores do artigo da ATEME, ainda argumentam que o sistema também pode abordar outros fatores-chave de configuração, afirmando que além de taxa, resolução, taxa de quadros e adaptação da faixa dinâmica, o framework proposto poderia até mesmo lidar com comutação de codec se necessário.

Para finalizar podemos deixar aqui uma provocação, dizendo que a melhor tecnologia de codificação por título seria aquela capaz de:

Poder mudar o número de degraus na escala de decodificação, assim como alterar a resolução dos degraus dessa mesma escala, enquanto oferece suporte para alterar as taxas de dados dos degraus na escala. Não podemos esquecer ainda a capacidade de dividir o vídeo em pequenas tomadas para codificação, ao em vez de usar GOPs. E já que estamos sonhando, a capacidade de poder utilizar vários codecs, com base nos dados recebidos pelos dispositivos dos usuários, também é algo desejável. Dessa forma poderíamos incorporar a suas métricas, analises de desempenho de rede, e ajustar a taxa de quadros, a faixa dinâmica e a gama de cores, baseado na capacidade desse desempenho medido. Não temos algo assim, ainda...mas espero voltar aqui para dar essa boa noticia em breve!

Referências e artigos citados:

1) https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2
2) https://beamr.com/cabr_library
3) https://www.researchgate.net/publication/333040039_Optimizing_Mass-Scale_Multi-Screen_Video_Delivery
4) https://mile-high.video/files/mhv2019/pdf/day1/1_06_Burnichon_Paper.pdf

O Autor:

Tom Jones Moreira

Responsável por implementar novas soluções para sistema de IPTV, ISDBT , DVBS e OTT. Coordenar equipes multidisciplinares na engenharia de Aplicação da Tecsys do Brasil. Desenvolvedor de papers técnicos científicos e painelista de eventos de tecnologia e novas mídias Membro do Fórum SBTVD, e FOBTV (Future of Broadcast TV).

Revisor Técnico da Revista da (SET)

Contato: tom@tecsysbrasil.com.br