

2 Conceitos Gerais do Padrão MPEG-2

Vários padrões de codificação, compactação e compressão foram desenvolvidos com o intuito de diminuir os requisitos de recursos para a transmissão e o armazenamento de sinais digitais de vídeo e áudio. Visando estabelecer padrões internacionais para a representação e codificação de informações audiovisuais em formato digital com compressão, a ISO (*International Organization for Standardization*) e a IEC (*International Electrotechnical Commission*) estabeleceram o grupo de trabalho MPEG (*Motion Picture Coding Experts Group*), que iniciou seus trabalhos em maio de 1988. A família de padrões produzidos foi popularmente conhecida como padrões MPEG e inclui, entre outros, os conjuntos de padrões MPEG1 e MPEG-2.

O padrão MPEG-2 foi iniciado em 1990, como uma evolução do MPEG1, e publicado em 1995. O objetivo deste padrão é prover uma taxa de vídeo de 1,5 Mbps a 15 Mbps, adequado para sinais de televisão padrão (SDTV – *Standard Definition Television*) e taxas de 15 Mbps a 30 Mbps para sinais de televisão de alta definição (HDTV – *High Definition Television*). Para taxas inferiores a 3 Mbps, o padrão MPEG1 pode apresentar maior eficiência que o MPEG-2.

O padrão MPEG-2 é descrito pelo conjunto de especificações ISO/IEC 13818, cujas principais definições estabelecem a forma de compressão para o fluxo multiplexado de sistemas [ISO00a], para o vídeo e para o áudio [ISO00b]. No âmbito do ITU-T (*International Telecommunication Union – Telecommunication Standardization Sector*), os padrões MPEG-2 *Systems* e MPEG-2 *Video* estão descritos nas recomendações H.222.0 [ITUT00a] e H.262 [ITUT00b], respectivamente.

Este capítulo descreve os principais conceitos do padrão MPEG-2, salientando os parâmetros que são diretamente alterados pelas operações de segmentação do vídeo em trechos e remontagem dos mesmos.

2.1. O Fluxo MPEG-2 de Sistemas

A estrutura de um fluxo definida pelo padrão MPEG-2 pode ser visualizada na Figura 2 e está dividida em duas camadas: a camada de compressão e a camada de sistema. A camada de sistema, definida no padrão *MPEG-2 Systems*, é responsável pela divisão e encapsulamento de cada fluxo comprimido em pacotes; pela inserção de informações de sincronização entre fluxos de mídias diferentes; pela multiplexação dos fluxos encapsulados; e pelo transporte da informação de referência do relógio utilizado no codificador. A camada de compressão refere-se à codificação de cada um dos dados audiovisuais, conforme especificado nos padrões MPEG-2 Áudio e Vídeo.

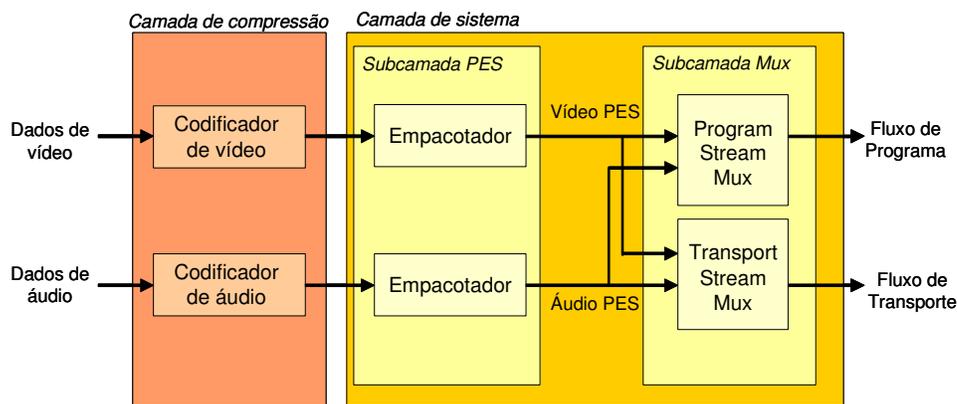


Figura 2 – Estrutura do fluxo MPEG-2

Os dados individuais de cada mídia, após sofrer o processo de compressão, são denominados de fluxos elementares e são divididos em pacotes na subcamada PES (*Packetized Elementary Stream*). As principais funções desempenhadas pela subcamada PES são a identificação exclusiva de cada fluxo, realizada através do parâmetro *stream_ID*, e a sincronização intra e intermídia, discutida a seguir. Os dados empacotados, ou seja, os fluxos de áudio, vídeo ou dados PES, são enviados à subcamada de multiplexação, onde é inserida a informação de referência de relógio do codificador.

Na camada de sistemas estão definidos dois formatos: o Fluxo de Transporte (TS), que contém um ou mais programas e é apropriado para a transmissão e o armazenamento em ambientes ruidosos, onde a ocorrência de erros é freqüente; e o Fluxo de Programa (PS), que contém apenas um programa e é adequado para uso em ambientes com baixas taxas de erros. Cada programa é definido como um

conjunto de fluxos elementares, vídeo, áudio e dados, por exemplo, que podem ou não ter algum relacionamento temporal entre si de um programa, utiliza uma mesma base de tempo, ou referência de relógio.

Através da inserção de marcas de tempo (*timestamp*), tanto nos fluxos PES quanto no fluxo de sistema, a sincronização intra e intermídia é obtida. A marca de tempo é uma amostra do contador da respectiva base de tempo, em um determinado instante. As marcas de tempo inseridas no fluxo de sistema, na subcamada de multiplexação, permitem, ao decodificador, a recuperação da referência do relógio utilizado pelo codificador. Elas são denominadas de *System Clock Reference* (SCR) e *Program Clock Reference* (PCR), respectivamente, para os fluxos TS e PS, e são definidas em termos de um relógio de sistema comum denominado *STC* (*System Time Clock*). Os valores das marcas de tempo SCR e PCR significam o instante de tempo em que o último bit desses campos entra no decodificador. O intervalo de tempo máximo permitido entre o envio de duas marcas consecutivas é de 0,7s. O processo de geração e extração das marcas de tempo relativas ao SCR e PCR é ilustrado na Figura 3. O padrão MPEG não considera os retardos introduzidos pela rede de comunicação.

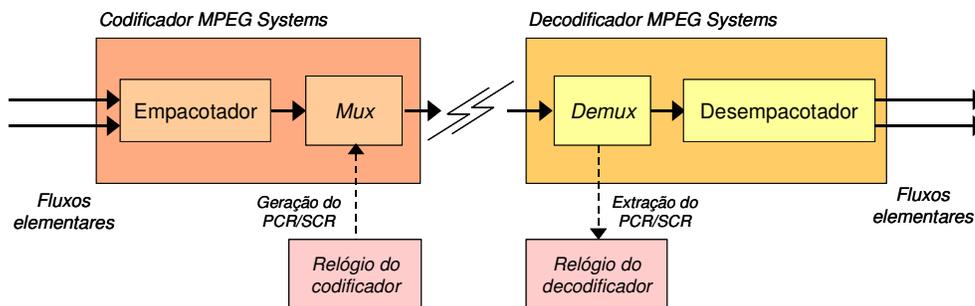


Figura 3 – Sincronização entre o codificador e o decodificador

Após o empacotamento dos respectivos dados em cada PES, alguns pacotes são escolhidos para transportar marcas de tempo. Dentre essas, dois tipos são definidos: o *Presentation Time Stamp* (PTS) e o *Decoding Time Stamp* (DTS). O PTS indica o instante de tempo em que a unidade de apresentação (figura, para o vídeo, e quadro, para o áudio) deve ser exibida. O DTS, presente apenas no fluxo de vídeo, indica o instante de tempo em que a unidade de apresentação deve ser entregue ao respectivo decodificador e é utilizado quando é necessária a reordenação de quadros, no decodificador.

2.2.
O Fluxo MPEG-2 de Vídeo

O padrão MPEG-2 de vídeo utiliza mecanismos para eliminar, ou reduzir, a redundância temporal existente entre quadros consecutivos. A estrutura do fluxo codificado através do MPEG-2 de vídeo é hierárquica e contém seis camadas: seqüência, grupo de quadros ou figura (GOP), quadro ou figura, *slice*, macrobloco e bloco, conforme ilustrado na Figura 4.

As imagens de um vídeo são representadas por quadros de vídeo, os quais são representados por três matrizes retangulares de inteiros: uma matriz de luminância e duas matrizes de crominância. O termo quadro é utilizado tanto para imagens ainda não codificadas através do padrão MPEG-2 quanto para as imagens após a codificação. As informações de um quadro podem ser separadas em campos denominados *top field* e *bottom field*, compostos por linhas ímpares e pares de cada matriz que compõe um quadro, respectivamente.

Um quadro codificado através do padrão MPEG-2 pode representar um quadro ou um campo codificado. Um sinal de vídeo que contenha quadros que representem campos é dito ser um vídeo entrelaçado. Se o fluxo de vídeo contiver apenas quadros que representem quadros, ele é dito ser progressivo.

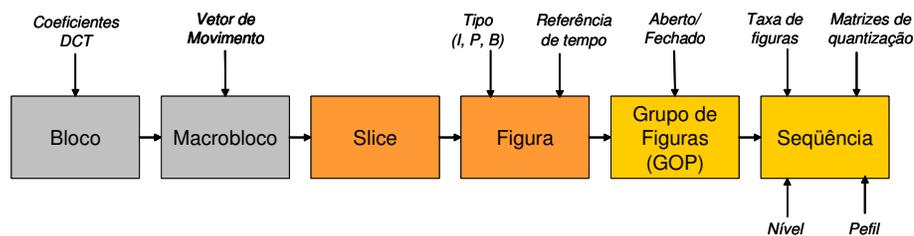


Figura 4 – Estrutura do fluxo MPEG Vídeo

2.2.1.
Estruturas de um fluxo de vídeo MPEG-2

A primeira camada do padrão MPEG-2 de vídeo codificado é a seqüência. Um fluxo de vídeo é composto por um conjunto de seqüências, que são utilizadas para transportar informações sobre as dimensões dos quadros e a sua relação de aspecto, as taxas de quadro e de bit, e as matrizes de quantização utilizadas na codificação dos macroblocos, caso os valores-padrão não sejam utilizados. O parâmetro *progressive_sequence*, presente no elemento *Sequence Extension*,

especifica o tipo de sinal de vídeo utilizado, ou seja, se o fluxo de vídeo contém quadros entrelaçados ou não.

A taxa de quadro é fixa, sendo determinada através dos parâmetros *frame_rate_code*, da estrutura *Sequence Header* e que indica o valor base (*frame_rate_value*) da taxa de quadros, e os parâmetros *frame_rate_extension_n* e *frame_rate_extension_d*, presentes na estrutura *Sequence Extension*. Ela é calculada conforme a equação abaixo:

$$\text{(Eq. 1) } frame_rate = frame_rate_value * \left[\frac{frame_rate_extension_n + 1}{frame_rate_extension_d + 1} \right]$$

A camada grupo de quadros ou figuras (GOP) informa o instante de tempo referente ao primeiro quadro que o compõe, de acordo com o valor do relógio de referência utilizado pelo codificador, e é utilizado para prover acesso aleatório ao fluxo de vídeo. Um GOP também indica a dependência temporal entre quadros pertencentes a GOP diferentes. GOP fechados são aqueles formados apenas por quadros cujas referências localizam-se no próprio GOP. Em um GOP aberto, a decodificação de um de seus quadros requer a decodificação anterior de um quadro localizado em outro GOP.

A camada de quadro ou figura especifica o tipo de predição utilizado na codificação de uma imagem e fornece a informação de temporização de cada quadro. A quantidade de bits utilizada para a codificação de cada quadro é variável e é influenciada pelo valor do *quantum* adotado no processo de quantização da imagem.

O método de predição interquadro por compensação de movimento realiza a comparação de cada macrobloco de um quadro com macroblocos de quadros vizinhos. O macrobloco do quadro vizinho, escolhido para servir de referência na operação de predição, será aquele que menos se diferenciar do macrobloco a ser codificado. Um vetor de movimento é definido para indicar a diferença entre as localizações espaciais do macrobloco a ser codificado e o de referência, sendo transmitido junto ao macrobloco codificado. Cada macrobloco especifica sua posição em relação ao macrobloco anterior, a indicação do método de predição utilizado e quais os blocos de luminância e crominância estão codificados.

A especificação da estrutura de um quadro, estabelecendo se esta representa um campo ou um quadro, é determinada, em um fluxo de vídeo codificado,

através do parâmetro *picture_structure*, presente no elemento *Picture Coding Extension* que está contido em cada quadro (elemento *Picture*). Para os propósitos desta dissertação, não serão considerados os quadros que representam campos. Os termos figura e quadro serão utilizados como sinônimos.

As imagens são codificadas em quadros do tipo I (*Intracoded*), P (*Predictive Coded*) ou B (*Bidirectional Predictive Coded*). Quadros I são codificados utilizando-se informações contidas no próprio quadro original. Quadros P são codificados de forma preditiva em relação ao quadro I ou P anterior. Por fim, quadros B são codificados de forma preditiva em relação aos quadros I ou P, anteriores ou posteriores. Dessa forma, para a decodificação de um quadro B, é necessário que o quadro posterior, ao qual aquele se referencia, já tenha sido decodificado.

Cada quadro codificado possui o parâmetro *temporal_reference* que funciona como um contador, módulo 1.024, o qual é incrementado a cada novo quadro e é utilizado para que o decodificador possa identificar eventuais perdas de quadros.

O parâmetro *vbv_delay*, contido no elemento *Picture Header*, indica o tempo que o quadro deve permanecer no *buffer* de entrada do decodificador, exceto quando possui valor hexadecimal FFFF.

Dois outros parâmetros, contidos no elemento *Picture Coding Extension*, são importantes para o processo de decodificação: *repeat_first_field* e *top_field_first*. Em vídeos progressivos, eles indicam a quantidade de vezes que uma figura deve ser apresentada, após sua decodificação. A relação entre esses parâmetros é mostrada na Tabela 1.

Tabela 1 – Parâmetros *repeat_first_field* e *top_field_first*

Parâmetros			Significado
<i>progressive sequence</i>	<i>Repeat first field</i>	<i>Top field first</i>	
1	0	0	Quadro deve ser apresentado uma vez
1	1	0	Quadro deve ser apresentado duas vezes
1	1	1	Quadro deve ser apresentado três vezes

A camada *slice* contém um conjunto de macroblocos, pertencentes a uma mesma linha da imagem codificada. No entanto, nem todos os macroblocos precisam ser inseridos no fluxo comprimido, os quais são chamados de *skipped macroblocks*. Em seu cabeçalho, um *slice* especifica a linha de macroblocos a que se refere e o fator de escala utilizado na determinação dos coeficientes DCT. Essa camada é especificada para facilitar o correto posicionamento espacial das amostras, no processo de exibição da imagem, em casos onde haja perda de dados. A perda de algumas amostras pode causar erro no posicionamento espacial das outras amostras pertencentes ao mesmo *slice*. Esse efeito, porém, não é cumulativo e os outros *slices* podem ser apresentados corretamente.

Cada macrobloco indica o modo de predição utilizado durante sua codificação e seu respectivo vetor de movimento. Também contém um fator de escala para permitir o controle do *quantum* do processo de quantização.

Um resumo das informações mais importantes definidas na sintaxe do padrão MPEG-2 de Vídeo e contidas em cada estrutura pode ser visualizado na Tabela 2.

Tabela 2 – Resumo da sintaxe das camadas do MPEG-2 de Vídeo

Nome da Camada	Elementos da sintaxe
Seqüência	Tamanho dos quadros
	Taxa de quadros por segundo
	Taxa de bits por segundo
	Tamanho do buffer de entrada do decodificador
	Parâmetros de codificação programáveis
GOP	Unidade de acesso aleatório
Quadro	Informação de temporização (temporal reference)
Slice	Informação de endereçamento em relação ao quadro
Macrobloco	Modo de codificação
	Vetores de movimento
	Quantização
Bloco	Coeficientes DCT

Para facilitar a decodificação, a ordenação das figuras no fluxo transmitido, ou ordem de codificação, é diferente da ordem segundo a qual as figuras devem ser exibidas, também chamada de ordem de apresentação. A ordem de codificação garante que os quadros codificados de forma preditiva são recebidos, pelo

decodificador, sempre após a recepção dos respectivos quadros utilizadas como referência. A Figura 5 exemplifica a disposição dos quadros segundo as ordens de apresentação e de codificação.

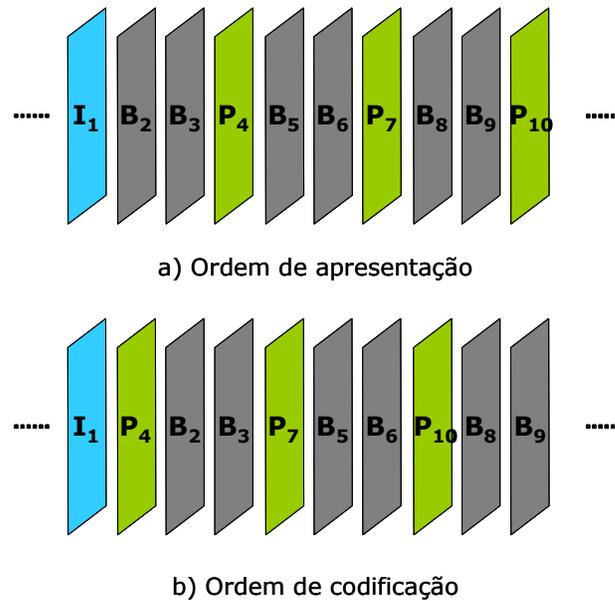


Figura 5 – Ordem de apresentação e codificação dos quadros.

2.2.2. Controle de ocupação do *buffer*

O gerenciamento da ocupação do *buffer* de entrada do receptor é realizado através da manipulação da quantidade de bits de cada figura, da taxa de bits do fluxo codificado e do tamanho do *buffer*, definido pelo codificador. O padrão MPEG-2 sugere a adoção do controle de taxa do modelo TM5 (*Test Model 5*), cujo objetivo é determinar, de forma adaptativa, o tipo de predição temporal para cada macrobloco, a matriz de quantização a ser aplicada e a taxa de transmissão, em bits por segundo, de cada quadro, de forma a evitar situações de *overflow* e *underflow* do *buffer* do receptor. As operações são realizadas nas camadas GOP, de quadro e de macrobloco, e são compostas por três atividades: a alocação de bits, o controle de taxa e a quantização adaptativa.

Na primeira etapa, um número fixo de bits é alocado para cada GOP, obtido através da taxa máxima de transmissão desejada, em bits por segundo, e o número de quadros contidos no GOP. Essa etapa é responsável pela estimativa do número de bits disponível para a codificação do próximo quadro, sendo realizada,

portanto, antes da codificação de cada quadro. Depois, à medida que cada quadro é codificado, a ocupação do *buffer* é monitorada, realimentando o sistema e definindo um valor de referência do *quantum* a ser utilizado por cada macrobloco. Por último, a quantização adaptativa manipula o valor de referência de acordo com a atividade espacial em cada macrobloco, de forma a determinar o valor exato do *quantum* relativo à codificação do respectivo macrobloco.

Detalhando o algoritmo especificado pelo TM5, inicialmente, é calculado o total de bits que o GOP deve conter, de acordo com a taxa de bits desejada, a taxa de quadros e o número de quadros dos tipos I, P e B que ainda devem ser inseridos no GOP corrente, conforme a equação abaixo:

$$\text{(Eq. 2)} \quad R_{GOP} = (N_i + N_p + N_b) \times \left(\frac{\text{bit_rate}}{\text{frame_rate}} \right)$$

Em seguida, são determinadas as “medidas de complexidade globais” (*global complexity measures*), denominadas de X_i , X_p ou X_b , de acordo com o tipo de quadro, segundo as seguintes fórmulas:

$$X_i = S_i Q_i$$

$$\text{(Eq. 3)} \quad X_p = S_p Q_p$$

$$X_b = S_b Q_b$$

Onde S_i , S_p e S_b representam o número de bits, conforme seu tipo, e as variáveis Q_i , Q_p e Q_b são os parâmetros de quantização médios, calculados de acordo com a média dos valores de quantização utilizados durante a codificação de todos os macroblocos. Todos os parâmetros anteriores referem-se aos últimos quadros codificados, de acordo com o respectivo tipo. Os valores iniciais para os parâmetros de quantização médios são:

$$X_i = \frac{(160 \times \text{bit_rate})}{115}$$

$$\text{(Eq. 4)} \quad X_p = \frac{(60 \times \text{bit_rate})}{115}$$

$$X_b = \frac{(42 \times \text{bit_rate})}{115}$$

A quantidade de bits (T_i , T_p e T_b) a ser utilizada na codificação do próxima quadro é determinada conforme as equações abaixo. Os parâmetros K_p e K_b são constantes cujos valores atribuídos são 1.0 e 1.4, respectivamente:

$$T_i = \max \left\{ \frac{R_{GOP}}{\left(1 + \frac{N_p X_p}{X_i X_p} + \frac{N_b X_b}{X_i K_b} \right)}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

$$(Eq. 5) T_p = \max \left\{ \frac{R_{GOP}}{\left(N_p + \frac{N_b K_p X_b}{K_b X_p} \right)}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

$$T_b = \max \left\{ \frac{R_{GOP}}{\left(N_b + \frac{N_p K_b X_p}{K_p X_b} \right)}, \frac{bit_rate}{8 \times picture_rate} \right\}$$

A segunda etapa refere-se ao controle da taxa e destina-se à obtenção do parâmetro de quantização Q_j referente ao macrobloco j , conforme a expressão:

$$(Eq. 6) Q_j = \left(\frac{d_j \times 31}{2} \right) \left(\frac{picture_rate}{bit_rate} \right)$$

onde d_j representa o nível de ocupação do *buffer* virtual e é calculado de acordo com o tipo de quadro.

A terceira etapa determina o valor de cada elemento da matriz de quantização a ser utilizado, a partir da atividade espacial (act_j) calculada para cada macrobloco e o respectivo parâmetro de quantização Q_j . O modelo TM5 especifica que:

$$(Eq. 7) act_j = 1 + \min(vblk_1, vblk_2, \dots, vblk_8)$$

$$(Eq. 8) vblk_n = \frac{1}{64} \times \sum_{k=1}^{64} (P_k^n - P_mean_n)^2$$

$$(Eq. 9) P_mean_n = \frac{1}{64} \times \sum_{k=1}^{64} P_k^n$$

O valor da atividade (act_j) é, então, normalizado, obtendo-se N_{act_j} e, finalmente, determinando-se o valor do *quantum* a ser aplicado ao respectivo macrobloco:

$$\text{(Eq. 10)} \quad mquant_j = Q_j \times N_{act_j}$$

Além do modelo TM5 do MPEG, outros algoritmos para o controle da ocupação do *buffer* do receptor foram propostos, tais como: o algoritmo de alocação de bits proposto por Song e Chun [SoCh03]; o esquema de controle de taxa através de histogramas baseado em estimativas de taxa e distorção (*Rate-Distortion Estimation*), proposto por Hong *et al.* [HYLK03]; e o algoritmo proposto por He e Mitra [HeMi02], que adota um relacionamento linear entre a taxa de codificação de bits e o percentual de zeros nos coeficientes DCT.

2.2.2.1. Situações de *overflow* e *underflow*

A ocorrência de *overflow* ou *underflow* do *buffer* de entrada do receptor é gerada por problemas no processo de codificação, alterando a taxa de produção de quadros em relação à taxa nominal. A taxa de codificação de quadros pode sofrer variações devido aos métodos utilizados para o cálculo da quantidade de bits e da taxa de transmissão dos dados, como, por exemplo, o estabelecido pelo TM5. O processo de codificação é responsável por, mesmo havendo pequenos desvios da taxa instantânea de codificação de quadros em relação à taxa nominal, manter a ocupação do *buffer* dentro dos limites adequados, evitando o *overflow* ou *underflow*. Fatores externos ao padrão MPEG também podem causar perturbações no decodificador, tais como os decorrentes da rede de comunicação.

A situação de *underflow* ocorre se a taxa de quadros gerada na saída do codificador for inferior à taxa de quadros prevista para o fluxo de vídeo. Nessa situação, o decodificador consome quadros do *buffer* mais rapidamente do que novos quadros entram no mesmo. Se essa situação persistir por um tempo suficientemente grande, todos os quadros do *buffer* serão retirados e o decodificador não encontrará um novo quadro completo para ser apresentado.

O *overflow* ocorre quando a taxa de entrada de quadros no *buffer* é superior à taxa nominal de apresentação de quadros, por um tempo suficientemente grande.

Os quadros se acumulam do *buffer* até que sua ocupação alcance o limite máximo. A partir desse momento, novos dados que cheguem ao *buffer* são descartados, gerando perda de quadros. Situações de *overflow* também podem acontecer se o algoritmo de alocação de bits, no processo de codificação, for inadequado e não limitar corretamente a quantidade de bits de cada figura.

Situações de *overflow* e *underflow* também podem ser provocadas por falhas na rede de comunicação devido a perdas de pacotes, fazendo com que a seqüência de quadros que entrem no *buffer* seja diferente da gerada pelo codificador. A perda de quadros, no entanto, prejudica a correta operação do decodificador, pois, caso tenham sido perdidos quadros configurados para terem sua apresentação repetida, o decodificador não recebe essa informação, adiantando a exibição do quadro seguinte. O funcionamento é normalizado após o recebimento de um novo GOP, que provê informações temporais para o acesso aleatório do fluxo.