

**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
Seção De Engenharia Elétrica / SE3**

1° Ten MÁRLISSON AUGUSTO ALVES MARQUES DA FONSECA

Transmissão de Vídeo sobre rede IP

Rio de Janeiro

2013

INSTITUTO MILITAR DE ENGENHARIA

1º Ten MÁRLISSON AUGUSTO ALVES MARQUES DA FONSECA

Transmissão de Vídeo sobre rede IP

Projeto de Fim de Curso apresentado ao Curso de Graduação em Engenharia Elétrica do Instituto Militar de Engenharia.

Orientadora: Carla Liberal Pagliari –Ph.D

Rio de Janeiro

2013

INSTITUTO MILITAR DE ENGENHARIA

1º Ten MÁRLISSON AUGUSTO ALVES MARQUES DA FONSECA

Transmissão de Vídeo sobre rede IP

Projeto de Fim de Curso apresentado ao Curso de Graduação em Engenharia Elétrica do Instituto Militar de Engenharia.

Orientadora: Carla Liberal Pagliari - Ph.D.

Aprovada em 03 de julho de 2013 pela seguinte Banca Examinadora:

Prof. Carla Liberal Pagliari - Ph.D., do IME

Major Mauro César Rebello Cordeiro - D.C., do IME

TC Alberto Gaspar Guimarães - D.C., do IME

Rio de Janeiro

2013

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do autor e do orientador.

623.043
A553e

Fonseca Márlisson Augusto Alves Marques da
Transmissão de Vídeo sobre rede IP/
Márlisson Augusto Alves Marques da Fonseca; orientado
por Carla Liberal Pagliari. - Rio de Janeiro: Instituto
Militar de Engenharia, 2013.

44 f.: Il., tab.

Monografia (Projeto Final de Curso) - Instituto Militar de
Engenharia – Rio de Janeiro, 2013.

1.Engenharias de Comunicações 2.H.264
3. rede IP I. Fonseca, Márlisson Augusto Alves
Marques da II. Pagliari, Carla Liberal III. Título V. Instituto
Militar de Engenharia

CDD 623.043

Sumário

1-INTRODUÇÃO	11
1.1- Justificativa	12
2- CONCEITOS BÁSICOS DE CODIFICAÇÃO DE VÍDEO	13
2.1- Padrão H.264/MPEG-4 AVC ¹	13
2.2- Codificador H.264.....	14
2.3- Predição	15
2.4- Transformada e Quantização	16
2.5- Codificação	16
2.6-Decodificação	17
2.7- Taxa de bits constante x Taxa de bits variável.....	17
2.8- Desempenho.....	18
3- ENCAPSULADOR MP4BOX	19
4- REDE UDP/IP	20
5- METODOLOGIA.....	22
5.1- Topologia e configuração da rede.....	22
5.2- Fluxo de vídeo simulado.....	24
5.3- A Estrutura Evalvid.....	28
6- RESULTADOS	29
7- CONCLUSÃO	43
8- REFERÊNCIAS BIBLIOGRÁFICAS.....	44

ÍNDICE DE FIGURAS

FIGURA 1 Diagrama em blocos do ambiente de simulação	12
FIGURA 2 Diagrama de blocos do codificador H.264 [4].....	15
FIGURA 3 O mesmo quadro, da esquerda para a direita, em vídeo bruto, em vídeo com codificação CBR, GOP IBBP e QP de frame "I" de 32 com perdas visíveis devido à quantização e vídeo codificado com QP de quadro "I" de 28, melhor qualidade de codificação, mas sofrendo os efeitos de uma perda de pacote.....	21
FIGURA 4 Diagrama em blocos do ambiente de simulação	22
FIGURA 5 Topologia física da rede adotada	23
FIGURA 6 Quadros 50, 150, 200 e 250 dos vídeos	24
FIGURA 7 Ordem de exibição e ordem de codificação	25
FIGURA 8 Perda de pacotes de tráfego de vídeo, em vermelho,.....	27
FIGURA 9 Frames I, B e P do vídeo Akiyo com codificação VBR, GOP IBP e QP 22	30
FIGURA 10 Frames I, B e P do vídeo Akiyo com codificação CBR, GOP IBP e QP 22	30
FIGURA 11 Frames I, B e P do vídeo Foreman com codificação VBR, GOP IBP e QP 22	31
FIGURA 12 Frames I, B e P do vídeo Foreman com codificação CBR, GOP IBP e QP 22	31
FIGURA 13 Frames I, B e P do vídeo Mobile com codificação VBR, GOP IBP e QP 22	32
FIGURA 14 Frames I, B e P do vídeo Mobile com codificação CBR, GOP IBP e QP 22.....	32
FIGURA 15 Gráfico Taxa x Distorção para os vídeos Akiyo, Foreman e Mobile com codificação VBR, GOP IBBP para os QP 22, 28 e 32, da esquerda para a direita	34
FIGURA 16 PSNR para diferentes codificações do vídeo Mobile	35
FIGURA 17 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	38
FIGURA 18 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	38
FIGURA 19 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	39
FIGURA 20 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	39
FIGURA 21 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	41
FIGURA 22 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo	41
FIGURA 23 Queda de PSNR devido a perdas de pacotes nas redes 3 e 4 do vídeo	42

ÍNDICE DE TABELAS

TABELA 1 Parâmetros variáveis das quatro redes adotadas.....	23
TABELA 2 Combinação de parâmetros possíveis à análise	27
TABELA 3 Quantidade média de pacotes em que foi dividido cada tipo de quadro para os diferentes parâmetros de codificação	35
TABELA 4 Taxas médias atingidas pelos vídeos Akiyo codificados	36
TABELA 5 Taxas médias atingidas pelos vídeos Foreman codificados	36
TABELA 6 Taxas médias atingidas pelos vídeos Mobile codificados	40

ABREVIACOES

AVC – *Advanced Video Coding*
CBR – *Constant Bit Rate*
CIF – *Common Intermediate Format*
DCT – *Discrete Cosine Transform*
GOP – *Group of Pictures*
IEC – *International Electrotechnical Commission*
IP – *Internet Protocol*
ISO – *International Organization for Standardization*
ITU – *International Telecommunication Union*
MB – *Macrobloco*
MPEG – *Moving Picture Experts Group*
MSE – *Mean Square Error*
NS2 – *Network Simulator 2*
PSNR – *Peak Signal-to-Noise Ratio*
QoS – *Quality of Service*
QP – *Quantization Parameter*
RTP – *Real-time Transport Protocol*
TCP – *Transmission Control Protocol*
UDP – *User Datagram Protocol*
VBR – *Variable Bit Rate*
VoIP – *Voice over Internet Protocol*

RESUMO

Este trabalho estuda o comportamento da transmissão de vídeos codificados com o padrão de compressão de H.264/MPEG-4 AVC¹ em redes IP. As redes são configuradas com diferentes parâmetros de banda-passante e fila, com finalidade de testar os efeitos nos pacotes das sequências de imagens. Os vídeos utilizados apresentam características diferentes em termos de textura e grau de movimento, visando explorar os limites das diferentes configurações de rede, incluindo tráfego de fundo competindo com o vídeo pela banda total.

Foram utilizadas ferramentas de codificação de vídeo, de encapsulamento e de simulação de redes.

ABSTRACT

This paper studies the behavior of encoded video streaming with the compression standard H.264/MPEG-4 AVC¹ in IP networks. The networks are configured with different bandpass and line parameters, with the purpose of testing the effects on packets of image sequences. The videos that were used have different characteristics in terms of texture and degree of movement in order to explore the limits of different network configurations, including background traffic competing with the video for the full band. Video encoding, encapsulation and network simulation tools were used.

1-INTRODUÇÃO

A transmissão de conteúdos multimídia em tempo-real através da Internet vem se tornando cada vez mais comum e consome mais recursos da rede mundial de computadores. Conseqüentemente, é cada vez maior a atenção dada à qualidade percebida pelos usuários, assim como é cada vez mais importante a preocupação com a influência que as aplicações multimídia em tempo-real sofrem e exercem sobre os demais fluxos que trafegam pela Internet. [5]

Dentre os conteúdos audiovisuais complexos que compõem os chamados fluxos multimídia, os fluxos de voz em banda estreita (limitada a aproximadamente 3,5kHz) e de vídeo codificado no padrão MPEG-4 Part 10 ou AVC (Advanced Video Coding)¹ já são bastante comuns na Internet, respectivamente na forma de serviços VoIP (*Voice over Internet Protocol* - voz sobre o protocolo da Internet) e serviços de Internet TV (televisão através da Internet).

No entanto, dentro do atual modelo de serviço *best effort* (melhor esforço) que a Internet oferece, não se podem dar garantias quanto a perdas e atrasos dos pacotes transmitidos, nem quanto à largura de banda disponível, porém pode-se procurar minimizar o impacto altamente nocivo que as perdas e os atrasos provocam na Qualidade do Serviço (QoS) de aplicações multimídia em tempo-real. Essas perdas normalmente são provocadas pelo descarte de pacotes em nós de rede que enfrentam congestionamento, onde o método de tratamento das filas é, na maioria das vezes, o *droptail*.

Assim, este trabalho avalia a influência em três exemplos de vídeos CIF, de mesmo tamanho mas com características de textura e movimento distintas. Os vídeos trafegam em quatro tipos de redes IP, com a mesma topologia física, mas diferentes combinações de bandas de transmissão e capacidade de *buffers* do tipo *droptail* nos nós, variando parâmetros de codificação, tais como: parâmetro de quantização (QP), grupo de imagens (GOP) e controle da taxa de bits (VBR ou CBR).

1.1- Justificativa

Com a crescente demanda de vídeos digitais em redes, faz-se necessário relacionar os principais tipos de falhas de vídeo recebido aos parâmetros de codificação e da rede IP. Assim, este trabalho visa investigar os efeitos no vídeo codificado e transmitido em diferentes cenários de redes de pacote.

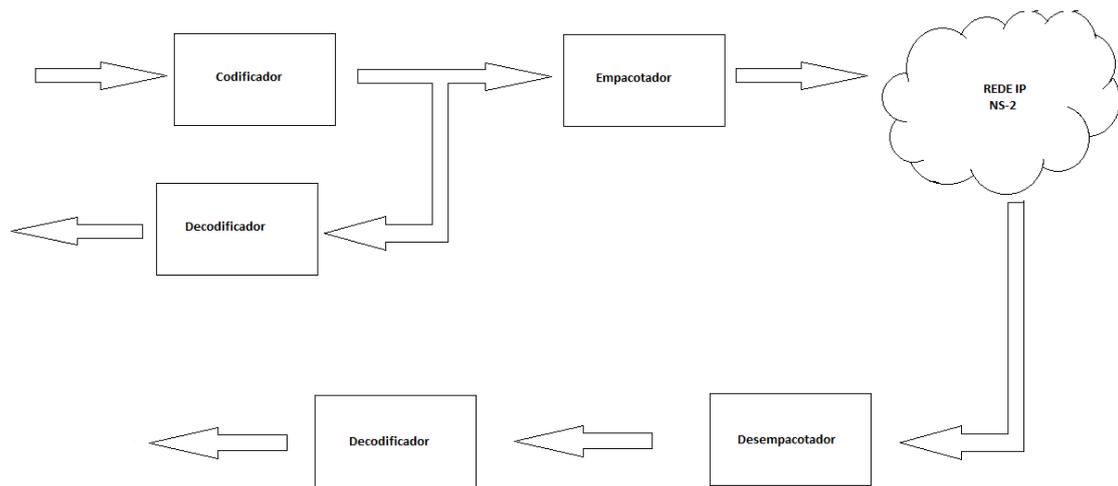


FIGURA 1 Diagrama em blocos do ambiente de simulação

2- CONCEITOS BÁSICOS DE CODIFICAÇÃO DE VÍDEO

2.1- Padrão H.264/MPEG-4 AVC¹

O H.264 Advanced Video Coding é um padrão da indústria para a codificação de vídeo. A compressão de vídeo ou de codificação de vídeo é uma tecnologia essencial para aplicações como TV digital, DVD-Video, TV móvel, videoconferência e transmissão de vídeo na Internet. A padronização de compressão de vídeo torna possível que produtos de diferentes fabricantes, como codificadores, decodificadores e meios de armazenamento tenham interoperabilidade.

A recomendação H.264 Advanced Video Coding é um documento copublicado por dois organismos internacionais de normalização, a ITU-T (International Telecommunication Union) e a ISO / IEC (International Organization for Standardization / International Electrotechnical Commission). Ele define um formato ou sintaxe de vídeo comprimido e um método para decodificar esta sintaxe para produzir uma sequência de vídeo que possa ser exibido. O documento padrão na verdade não especificar como codificar vídeo digital - isso fica a cargo do fabricante de um codificador de vídeo - mas na prática é provável que o codificador espelhe as etapas do processo de decodificação.

O padrão H.264/AVC foi publicado pela primeira vez em 2003, com várias revisões e atualizações publicadas desde então. Baseia-se nos conceitos de normas anteriores, como MPEG-2 e MPEG-4 Visual e oferece o potencial para uma melhor eficiência de compressão, ou seja, melhor qualidade de vídeo comprimido, e uma maior flexibilidade na compressão, transmissão e armazenamento de vídeo.

H.264/AVC descreve um conjunto de ferramentas ou métodos de compressão de vídeo. A norma especifica como o vídeo codificado com essas ferramentas deve ser representado e decodificado. Um codificador de vídeo pode escolher quais ferramentas utilizar e como aplicá-las para a sequência de vídeo atual.

H.264 oferece um desempenho significativamente melhor do que a compressão MPEG-2. Usando H.264 é possível comprimir o vídeo num número muito menor de bits do que usando o MPEG-2, para a mesma resolução e qualidade de imagem de vídeo. Isto significa, por exemplo, que muito mais material de vídeo pode ser armazenado num disco ou transmitido através de um canal de transmissão, utilizando o formato H.264.

2.2- Codificador H.264

Quando uma sequência de quadros de vídeo originais é codificado para o formato H.264, uma série de bits representa o vídeo em forma comprimida. Este fluxo contínuo de dados comprimido é armazenado ou transmitido e pode ser decodificado para reconstituir a sequência de vídeo. A versão decodificada, em geral, não é idêntica à sequência original, porque o H.264 é um formato de compressão com perdas, isto é, uma qualidade de imagem é perdida durante a compressão.

Os dados são processados em unidades de um macrobloco (MB), geralmente correspondente a 16×16 pixels. No codificador, um macrobloco de predição é gerado e subtraído do macrobloco atual para formar um macrobloco residual, o que é transformado, quantizado e codificado. Em paralelo, os dados quantizados são submetidos à transformada inversa e adicionado à predição do macrobloco para reconstruir uma versão codificada do quadro que é armazenada para posteriores predições. No decodificador, um macrobloco é decodificado, reordenado e transformado inversamente de modo a formar um macrobloco residual decodificado. O decodificador gera a mesma previsão que foi criado no codificador e adiciona este ao resíduo para produzir um macrobloco decodificado.

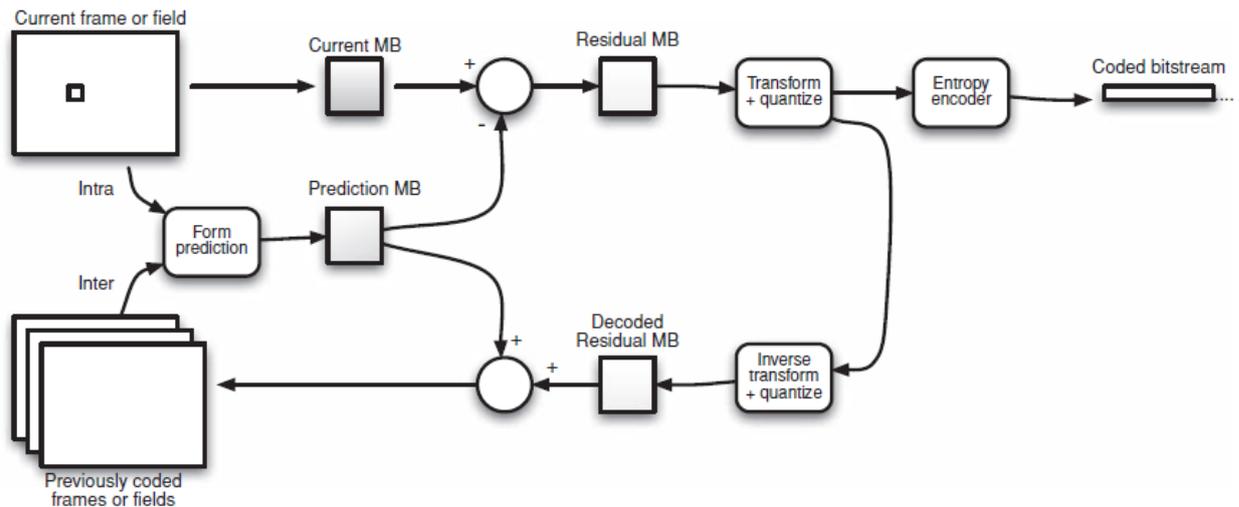


FIGURA 2 Diagrama de blocos do codificador H.264 [4]

2.3- Predição

O codificador forma uma predição do macrobloco atual, baseado em dados previamente codificados, quer a partir do quadro atual utilizando previsão intra ou de outros quadros que já foram codificadas e transmitidas utilizando previsão Inter. O codificador subtrai a predição do macrobloco atual para formar um resíduo.

Os métodos de previsão suportados pelo H.264 são mais flexíveis do que os padrões anteriores, permitindo previsões precisas e com isso compressão de vídeo mais eficiente. Na predição Intra, os valores dos pixels vizinhos anteriormente codificados são extrapolados para formar uma predição do macrobloco atual.

A predição Inter utiliza uma gama de tamanhos de bloco de 16 x 16 até 4 x 4 para prever pixels no quadro atual das regiões semelhantes em quadros codificados anteriormente. Esses quadros anteriormente codificados podem ocorrer antes ou após o quadro atual na ordem de exibição.

2.4- Transformada e Quantização

Um bloco de amostras residuais é transformado usando transformações aproximadas da transformação de cosseno discreta (DCT). A transformada produz um conjunto de coeficientes, cada um dos quais é um valor de peso para um padrão de base normalizada. Quando combinados, os padrões de base ponderada recriam o bloco de amostras residuais.

A saída da transformada, de um bloco de coeficientes transformados é quantizada. A quantização reduz a precisão dos coeficientes de transformada de acordo com um parâmetro (QP). Definir QP para um valor alto resulta em alta compressão à custa da má qualidade de imagem decodificada. Definindo QP para um valor mais baixo significa que os coeficientes não zero permanecem após a quantização, resultando em melhor qualidade da imagem no decodificador, mas também na de compressão mais baixa.

2.5- Codificação

O processo de codificação de vídeo produz uma série de valores que devem ser codificados para formar o fluxo de dados comprimidos. Estes valores incluem:

- Coeficientes da transformação quantizados
- Informações que permitam o decodificador para recriar a previsão
- Informação sobre a estrutura dos dados comprimidos e as ferramentas de compressão utilizadas durante a codificação
- Informações sobre a sequência de vídeo completa.

Estes valores e parâmetros, elementos de sintaxe, são convertidos em códigos binários usando codificação de comprimento variável e/ou codificação aritmética. Cada um destes métodos de codificação produz uma representação binária eficiente e compacta da informação. A corrente de bits codificada, em seguida, pode ser armazenada e/ou transmitida.

2.6-Decodificação

Um decodificador de vídeo recebe o fluxo de bits H.264 comprimido, decodifica cada um dos elementos de sintaxe e extrai as informações descritas acima, ou seja, transformar os coeficientes quantizados, informações de previsão, etc. Esta informação é então usada para reverter o processo de codificação e recriar uma sequência de vídeo.

Os coeficientes quantizados são redimensionados. Cada coeficiente é multiplicado por um valor inteiro para restaurar a sua escala original, de acordo com o QP adotado. Os coeficientes reescalados são semelhantes, mas não idênticas aos originais. A diferença é a perda devido ao processo de quantificação. O parâmetro de quantização maior tende a produzir uma maior diferença entre os blocos originais e reconstruídos.

Para cada macrobloco, o decodificador constitui uma previsão idêntica à que foi criada pelo codificador utilizando predição intra de quadros previamente decodificados ou predição intra a partir de amostras previamente decodificados no quadro atual. O decodificador adiciona a previsão no decodificador residual para reconstituir um macrobloco decodificado, que pode então ser apresentada como parte de um quadro de vídeo.

2.7- Taxa de bits constante x Taxa de bits variável

Existem duas maneiras principais para se codificar sinais de vídeo:

- a) com escalas de quantização fixas, resultando em qualidade praticamente constante às custas de uma taxa de bits variável;
- b) com controle de taxa, que adapta as escalas de quantização de forma a manter a taxa de transmissão praticamente constante à custa de uma qualidade variável.

Nas codificações com controle de taxa realizadas neste trabalho apenas os frames tipo “P” e “B” têm a liberdade de variar a escala de quantização.

2.8- Desempenho

A maior vantagem do H.264 sobre padrões anteriores é o seu desempenho de compressão. Comparado com os padrões como o MPEG-2 e MPEG-4 Visual, o H.264 pode entregar:

- Melhor qualidade de imagem ao mesmo bitrate de comprimido, ou
- Uma taxa de bits de comprimido inferior para a mesma qualidade da imagem.

Em termos comparativos, o padrão H.264/MPEG-4 AVC atinge o mesmo nível de qualidade percebida utilizando aproximadamente metade da banda (ou, dependendo da aplicação, metade do espaço de armazenamento) que o antigo padrão MPEG-2, usado nos DVDs e nos primeiros sistemas de TV digital, por exemplo. Visto por outro ângulo, uma mesma rede de acesso consegue fornecer serviços de vídeo com muito mais qualidade, se comparado ao padrão MPEG-2, graças à capacidade de compressão do H.264/MPEG-4 AVC.

Por exemplo, um DVD de camada única pode armazenar um filme de cerca de duas horas de duração no formato MPEG-2. Usando H.264, deve ser possível armazenar quatro horas ou mais de vídeo com qualidade de filme no mesmo disco, isto é, uma taxa de bits mais baixa para a mesma qualidade. Alternativamente, o formato de compressão H.264 pode oferecer uma melhor qualidade na mesma taxa de bits em comparação com MPEG-2 e MPEG-4 Visual.

3- ENCAPSULADOR MP4BOX

MP4Box é um empacotador multimídia, que encapsula *streams* de vídeo em formato ISO MP4, de forma a ser possível empacotá-los e enviar por RTP (*Real-time Transport Protocol*). O RTP é um protocolo de redes utilizado em aplicações de tempo real. Ele define como deve ser feita a fragmentação do fluxo de dados, adicionando a cada fragmento informação de sequência e tempo de entrega. O RTP usa o UDP como protocolo de transporte, o qual não oferece qualquer garantia que os pacotes serão entregues num determinado intervalo de tempo.

4- REDE UDP/IP

A Internet pode ser definida como um conjunto de diversas redes interconectadas que utilizam o protocolo IP na camada de rede para a transmissão de pacotes de informação.

Como uma parcela significativa das aplicações multimídia presentes na Internet, tais como VoIP, fluxos de áudio e vídeo em tempo-real e videoconferências, possuem características conversacionais e restrições muito fortes em termos de tempo de transmissão e exigência de largura de banda, tornou-se necessária a utilização de uma combinação alternativa ao TCP/IP. [3].

Daí surgiu a combinação entre o UDP e o IP, onde a confiabilidade e o “bom comportamento” do TCP dão lugar à simplicidade do UDP, que apenas envia mensagens com um mínimo de mecanismos envolvidos, sem garantia de entrega nem proteção contra entregas duplicadas. Basicamente, o que o UDP faz é, do lado do transmissor, repassar ao IP segmentos de dados de uma determinada aplicação para que este faça o roteamento até o destino e, do lado do receptor, repassar os segmentos do IP para a aplicação. Com o UDP, se forem necessários esquemas adicionais de controle de erro terão de ser implementados numa camada superior.

Na transmissão de vídeo comprimido através de redes IP, a degradação da qualidade pode ocorrer devido aos efeitos inerentes ao processo de codificação, como por exemplo o efeito de bordas borradas, bordas quadriculadas, etc. Quando o vídeo é transmitido via rede pode ocorrer perda de pacotes e o decodificador não vai receber toda a informação necessária para reconstruir o vídeo transmitido. Os efeitos, em geral, são perceptualmente bem mais acentuados. Muitos reprodutores de mídia (*media player*) possuem ferramentas de “*error concealment*” que tentam amenizar os efeitos de perda de pacotes, ilustrado na FIGURA 3.



FIGURA 3 O mesmo quadro, da esquerda para a direita, em vídeo bruto, em vídeo com codificação CBR, GOP IBBP e QP de frame "I" de 32 com perdas visíveis devido à quantização e vídeo codificado com QP de quadro "I" de 28, melhor qualidade de codificação, mas sofrendo os efeitos de uma perda de pacote.

5- METODOLOGIA

A metodologia escolhida para este trabalho foi a simulação computacional de uma rede IP que transmite um fluxo de vídeo H.264/AVC empacotado pela ferramenta MP4Box, em tempo real, disputando os recursos da rede com um fluxo de bits constante.

Os vídeos foram codificados offline usando o software de referência JM versão 18.4 [8] e encapsulados usando o software MP4Box. As simulações foram implementadas no software de simulação de redes Network Simulator versão 2 (NS-2) [6] e a análise de dados foi feita offline a partir dos arquivos trace gerados pelo NS-2.

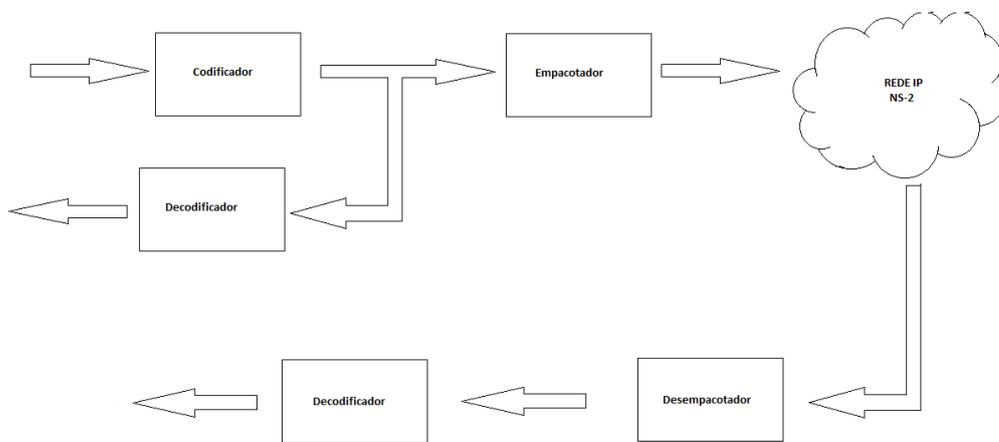


FIGURA 4 Diagrama em blocos do ambiente de simulação

5.1- Topologia e configuração da rede

Na topologia da rede utilizada, os roteadores R2 e R3 são os nós intermediários entre o enlace de tráfego concorrente, oriundos do tráfego de vídeo entre a fonte Tx1 e o destino Rx5, e o enlace em que trafega um fluxo constante de bits (CBR), entre a fonte de bits a taxa constante Tx0 ao destino Rx4, simulando outro tráfego UDP concorrente. Foram simulados quatro exemplos de rede com a mesma topologia física apenas diferenciando entre si na banda passante dos ramos e no tamanho de *buffer*. Nos demais fatores, os enlaces tiveram a mesma configuração de tempo de retardo de 32 ms .

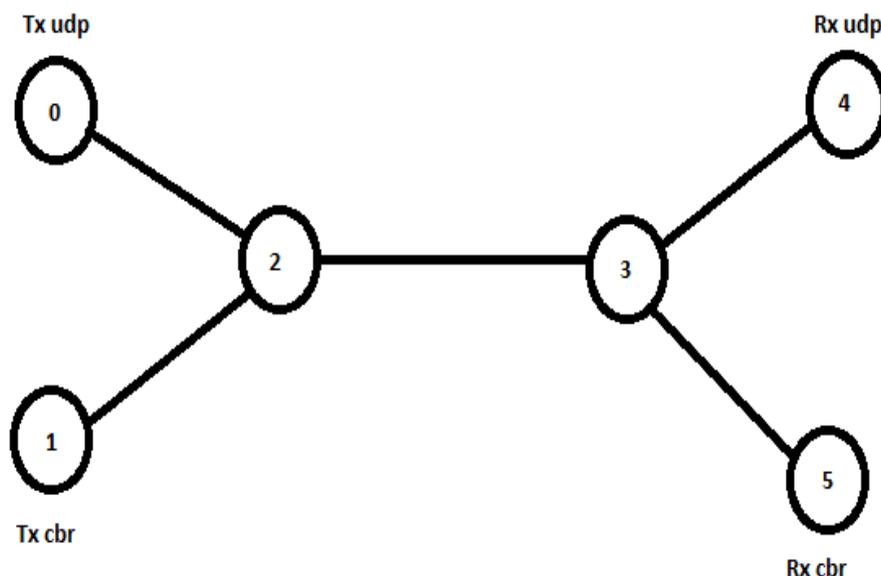


FIGURA 5 Topologia física da rede adotada

Os trafego de vídeo é disparado no instante 0 e tem duração de 10 segundos, com taxa de 30 quadros por segundo, resultando num total de 300 quadros. O fluxo concorrente de bits a taxa constante é iniciado no mesmo instante, porém encerrado no instante de 9.5 segundos, a uma taxa constante de 518,4 kbps. Foi também configurado o tamanho máximo de pacotes do na rede como 1028 bytes, sendo 1000 bytes o tamanho máximo dos pacotes encapsulados no MP4Box, 20 bytes do protocolo IP e 8 bytes de header UDP.

As quatro configurações de rede utilizadas seguem na TABELA 1:

	Taxa Tx0-R2 (Mbps)	Taxa Tx1-R2 (Mbps)	Taxa R2-R3 (Mbps)	Taxa R2-Rx4 (Mbps)	Taxa R2-Rx5 (Mbps)	Tamanho do <i>buffer</i> (pacotes)
Rede 1	1.1	1.1	1.4	1.1	1.1	40
Rede 2	1.1	1.1	1.4	1.1	1.1	80
Rede 3	0.9	0.9	1.0	0.9	0.9	40
Rede 4	0.9	0.9	1.0	0.9	0.9	80

TABELA 1 Parâmetros variáveis das quatro redes adotadas

5.2- Fluxo de vídeo simulado

Os fluxos de vídeo H.264/MPEG-4 AVC foram gerados a partir de sequências de vídeo bruto no formato CIF 352x288 pixels, 4:2:0 YUV, com 10 segundos de duração e amostrados a 30 quadros por segundo. As sequências utilizadas foram Akiyo.yuv, Foreman.yuv e Mobile.yuv, todas disponíveis em <<http://trace.eas.asu.edu/yuv/index.html>>.

A FIGURA 6 mostra os quadros de número 50, 100, 150, 200 e 250 (de um total de 300 quadros) dos vídeos “Akiyo”, “Foreman” e “Mobile”, a fim de demonstrar a diferença em temas de variação de quadros entre as sequências: a primeira sequência em que predominam semelhanças entre quadros, com uma apresentadora de telejornal praticamente estáticos e um fundo que varia muito pouco (Akiyo), a última em que na sequência predominam diferenças entre quadros, com o trenzinho, o calendário, outros objetos e o fundo da imagem em constante movimento (Mobile) e o segundo, em com características intermediárias dos anteriores (Foreman).



FIGURA 6 Quadros 50, 150, 200 e 250 dos vídeos Akiyo, Foreman e Mobile, respectivamente

As codificações foram feitas através do software de referência JM versão 18.4, usando o perfil “High”, que suporta vídeo amostrado em 8 bits por amostra no formato 4:2:0. Os GOPs (*Group of Pictures* – grupo de imagens) utilizados foram dos tipos IBP (período 12) e IBBP (período 15), onde os quadros intra-codificado (tipo I) e os quadros inter-codificados (tipo P) são intercalados por um, no primeiro caso, ou dois, no segundo, quadros codificados bi-direcionalmente (tipo B).

Pode-se observar que geralmente os quadros do tipo I tem tamanhos maiores, e podem ser codificados/decodificados independentemente. Os quadros do tipo I são representados com mais bits, pois exploram somente a redundância espacial. Já os do tipo P e B, que são estimados via predição temporal, são representados com um número inferior de bits. Os do tipo B tem, em geral, tem tamanho inferior aos quadros do tipo P já que empregam estimação de movimento bidirecional explorando melhor a predição de movimento. A FIGURA 7 apresenta a estrutura temporal IBBP.

É importante distinguir ordens diferentes de imagens: ordem de codificação, a ordem em que as imagens são codificados a partir do fluxo de bits, e ordem de exibição.

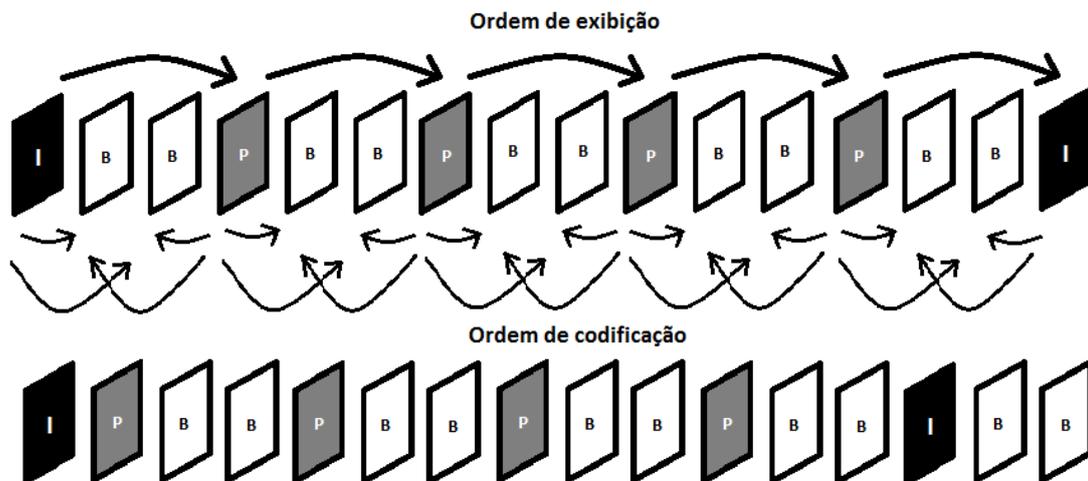


FIGURA 7 Ordem de exibição e ordem de codificação dos quadros com GOP IBBP de período 15

Além disso, outra variação de configuração de codificação foi utilizada: com controle de fluxo H.264/MPEG-4 AVC desabilitado e habilitado. Dessa forma, a saída do codificador resulta num fluxo de vídeo do tipo VBR (*Variable Bit Rate* – Taxa de Bits Variável), ou num tipo CBR (*Constant Bit Rate* – Taxa de Bits Constante). Nas codificações com controle de taxa realizadas neste trabalho apenas os frames tipo “P” e “B” têm a liberdade de variar a escala de quantização.

Uma vez terminada a codificação, o vídeo em formato H.264 foi encapsulado para *containers* MP4 utilizando a ferramenta MP4Box, disponível em <media.xiph.org/vídeo/derf>. Nessa etapa, cada quadro foi segmentado em pacotes de 1028 bytes.

Utilizando a ferramenta MP4TRACE, ferramenta do Evalvid foram gerados os arquivos *trace* usados na simulação. A partir dos arquivos *trace* de saída do NS-2 foram calculadas as perdas de pacotes de vídeo e a distorção sofrida após terem sido transmitidos pela rede.

A métrica usada para medir o grau de distorção é o PSNR (Peak Signal Noise Ratio), dado pela fórmula:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

Onde MAX é o valor máximo possível de um pixel e MSE (Mean Square Error):

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i,j) - K(i,j)\|^2$$

A FIGURA 8 exibe um *screenshot* do NAM, parte gráfica do NS-2, em uma simulação. As linhas vermelhas representam o fluxo de bits de vídeo do nó 1 ao nó 5, as linhas azuis o fluxo de fundo do nó 0 ao nó 4. Os quadrados empilhados acima de cada nó representa a fila de pacotes do *buffer* e os quadrados “caindo” representam as perdas de pacote devido à saturação do *buffer*.

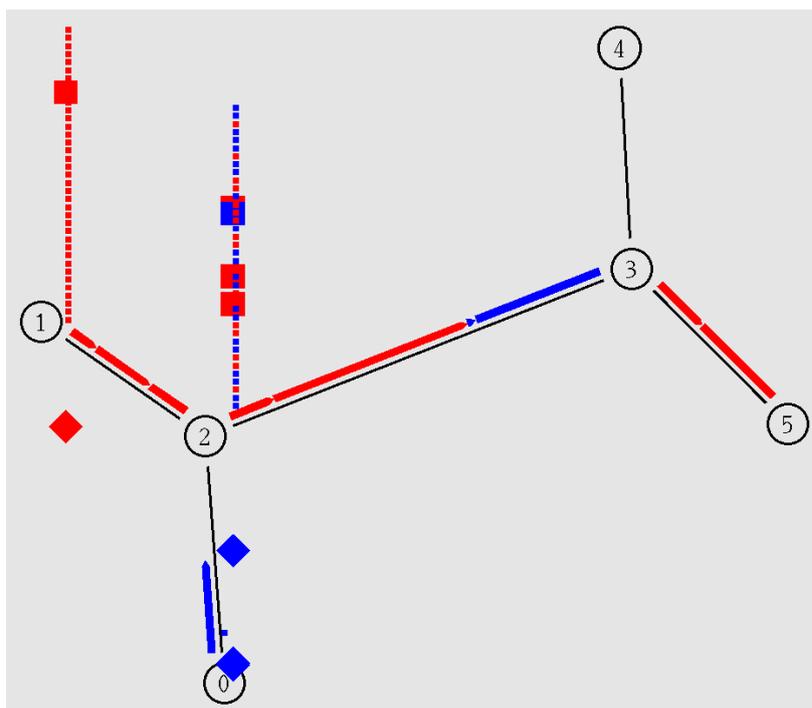


FIGURA 8 Perda de pacotes de tráfego de vídeo, em vermelho, e de tráfego de fundo, em azul

Os parâmetros de codificação estão resumidos na TABELA 2:

PARÂMETROS	DE CODIFICAÇÃO			DE REDE	
	GOP	Controle de Fluxo	Passo de Quantização	Banda Passante	Tamanho de Buffer
Opções	IBBP e IBP	VBR e CBR	22, 28 e 32	1.1/1.4 e 0.9/1.0	40 e 80
Nº de opções	2	2	3	2	2

TABELA 2 Combinação de parâmetros possíveis à análise

No total, temos 12 configurações diferentes de codificação, e 4 tipos diferentes de rede (1, 2, 3 e 4). Só então, temos 48 tipos de análise que podem ser feitas. Como foram simulados a codificação e a passagem de na rede IP foi simulada para 3 tipos diferentes de vídeos, obteve-se 144 resultados.

As bandas passantes escolhidas são compatíveis com as taxas médias geradas pela codificação de vídeos CIF. Os *buffers* apresentam fila do tipo *DropTail* que são as mais simples, é um tipo FIFO (first-in-first-out), e comuns de serem aplicadas em tráfego em Internet.

5.3- A Estrutura Evalvid

Evalvid [7] é o nome de uma estrutura formada por um conjunto de ferramentas para avaliação da qualidade de fluxos de vídeo transmitidos em redes reais ou simuladas. O que torna a estrutura “Evalvid” atraente é a relativa simplicidade da interação entre os arquivos de *trace* usados na avaliação da qualidade de transmissão: são necessários apenas os arquivos de *trace* de entrada (representando o fluxo de vídeo original) e de saída da rede (representando o fluxo de vídeo corrompido). Dessa forma, a rede pode ser encarada como uma “caixa preta” que provoca perdas, atrasos e um possível reordenamento dos pacotes. Assim como a “caixa” que representa a rede, as demais “caixas” que formam a estrutura também podem ser substituídas. Essa característica é particularmente interessante para o caso da “caixa” que representa o codificador de vídeo.

Através da ferramenta MP4TRACE, as sequências de vídeo codificadas e preparadas para transmissão foram usadas como entrada para a geração de um arquivo de texto do tipo *trace*, que contém todas as informações necessárias para que o NS2 crie um fluxo simulado que se comporta exatamente como o fluxo codificado original em termos de número de sequência de quadros, tipo de quadro, quantidade e tamanho dos segmentos de quadros (pacotes) enviados, e temporização de transmissão. Os valores de PSNR foram calculados através da ferramenta PSNR do Evalvid a partir dos vídeos reconstruídos do lado do decodificador, último bloco na sequência representado na FIGURA 4.

6- RESULTADOS

Com as diferentes configurações de codificação na TABELA 2, pôde-se observar que no geral os quadros do tipo I são representados com mais bits, pois somente exploram a redundância espacial. Já os do tipo P e B, que são estimados via predição temporal, são representados com um número inferior de bits. Os do tipo B tem, em geral, tamanho inferior aos quadros do tipo P já que empregam estimação de movimento bi-direcional explorando melhor a predição de movimento

As FIGURAS 9, 10, 11, 12, 13 e 14 representam os quadros da codificação do vídeo Akiyo, Foreman e MóBILE com as diferentes configurações de codificação. As barras vermelhas representam os quadros tipo I, as azuis os quadros tipo P e as verdes os quadros tipo B.

Neste primeiro momento, serão analisados apenas os gráficos de codificações VBR, ou seja, aquelas em que o parâmetro de quantização foi mantido constante enquanto a taxa, e tamanho dos frames, pode variar livremente.

A notação na codificação CBR acompanhada de valor numérico, como em CBR IBP 22, não representa que todos os frames foram codificados com QP constante, apenas os frames tipo “I”. Desta forma, foi adotado neste trabalho que é a codificação dos frames “P” e “B” tem a incumbência de ajustar seu QP para tentar alcançar a taxa alvo.

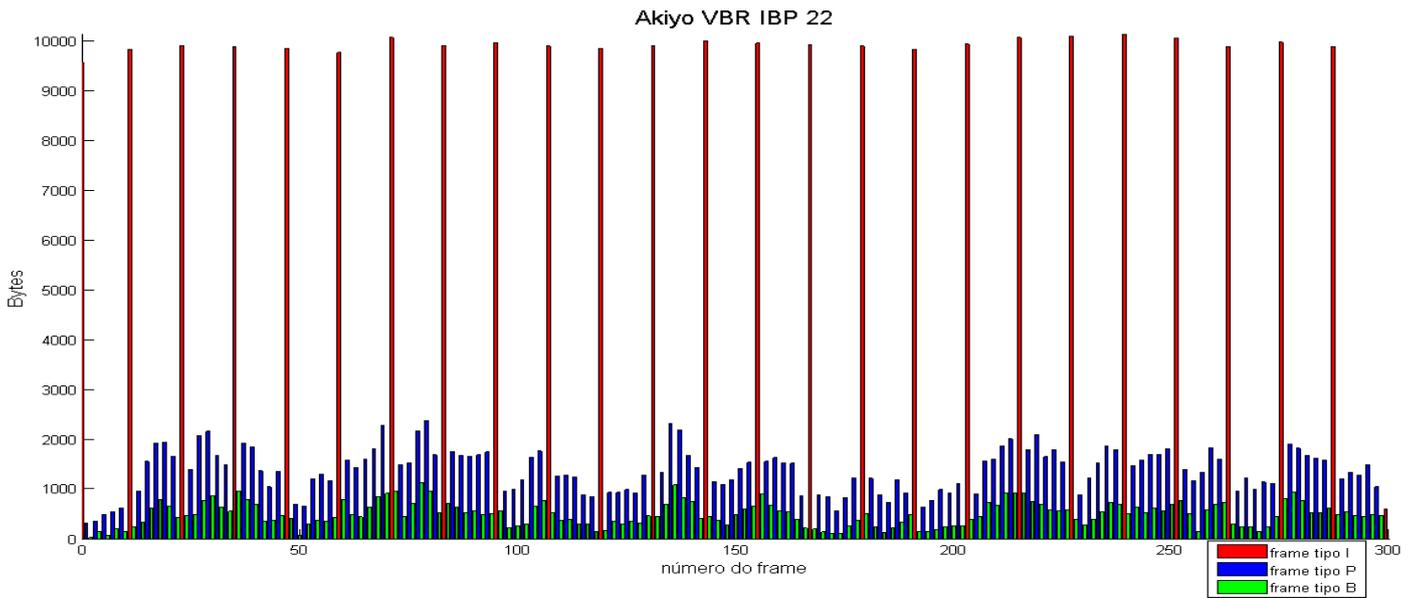


FIGURA 9 Frames I, B e P do vídeo Akiyo com codificação VBR, GOP IBP e QP 22

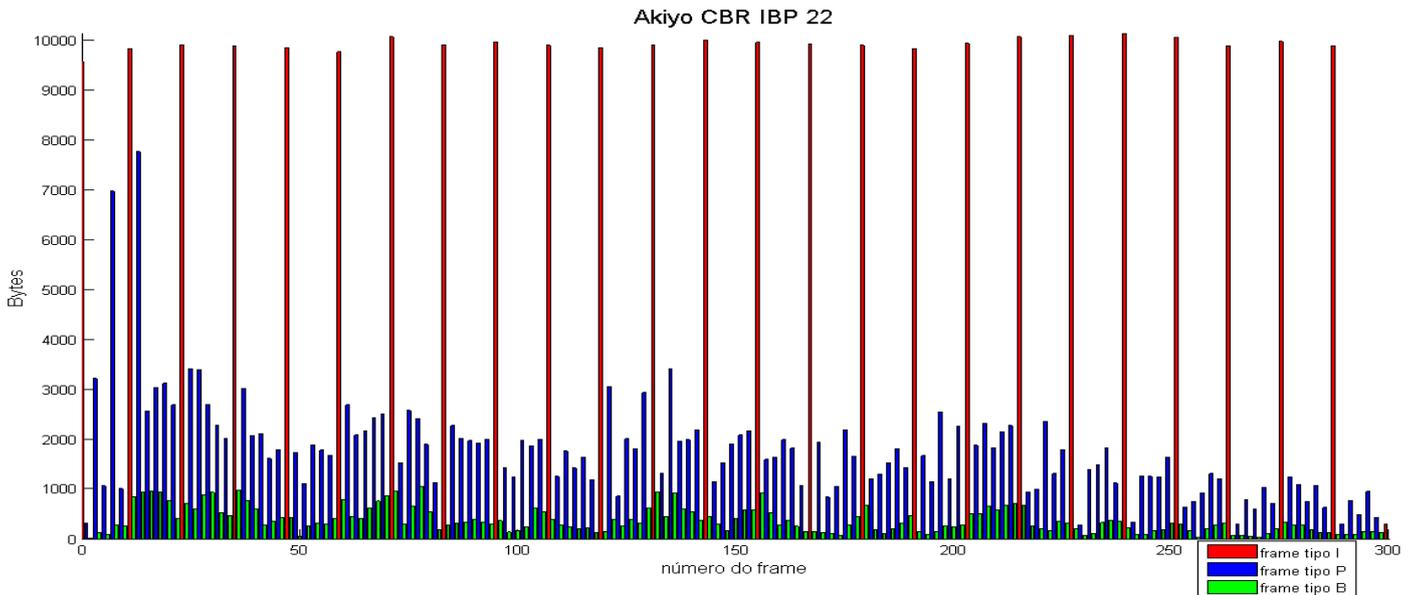


FIGURA 10 Frames I, B e P do vídeo Akiyo com codificação CBR, GOP IBP e QP 22

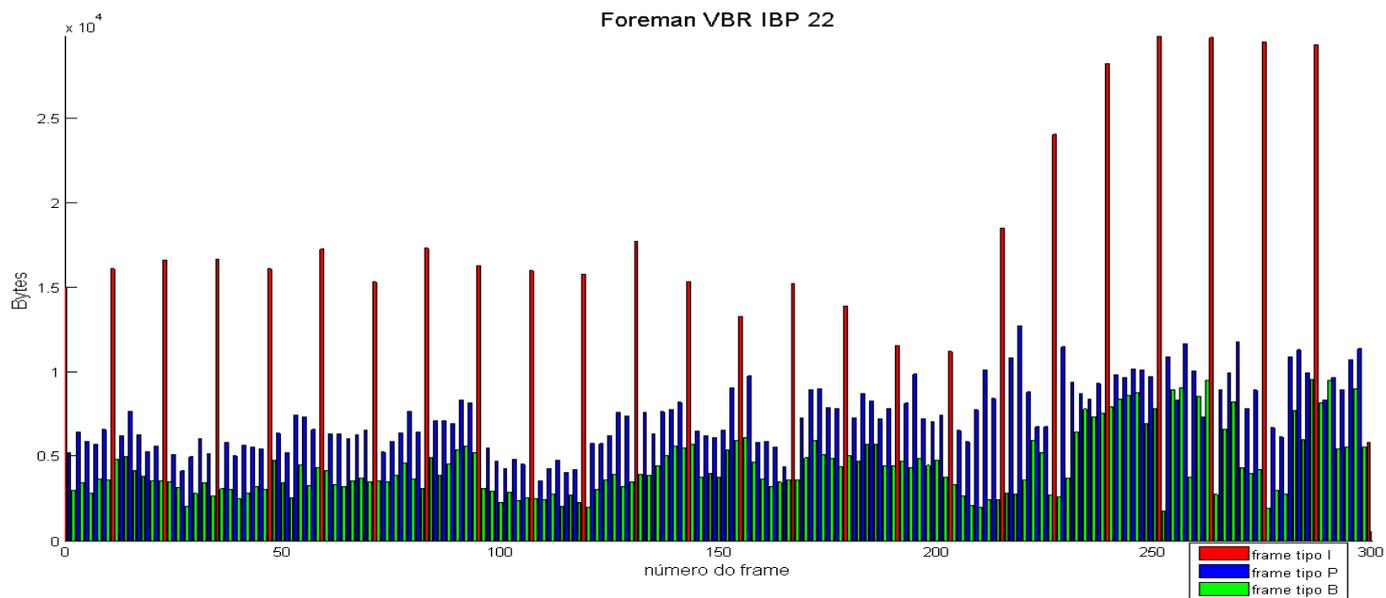


FIGURA 11 Frames I, B e P do vídeo Foreman com codificação VBR, GOP IBP e QP 22

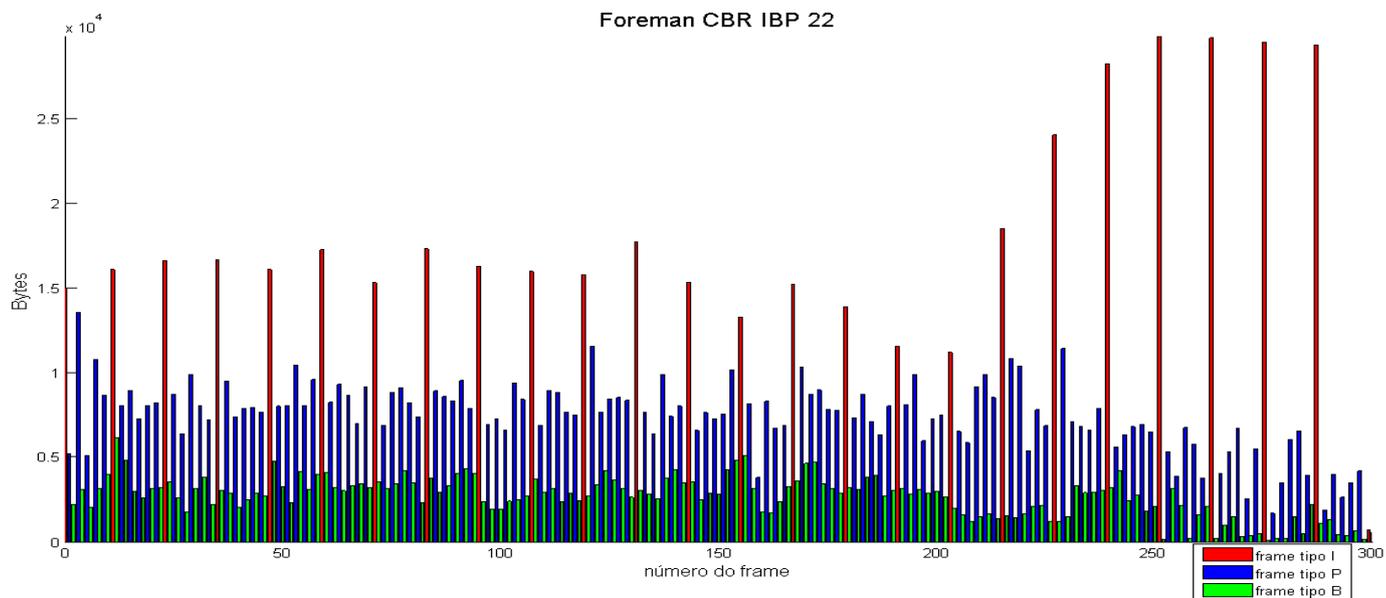


FIGURA 12 Frames I, B e P do vídeo Foreman com codificação CBR, GOP IBP e QP 22

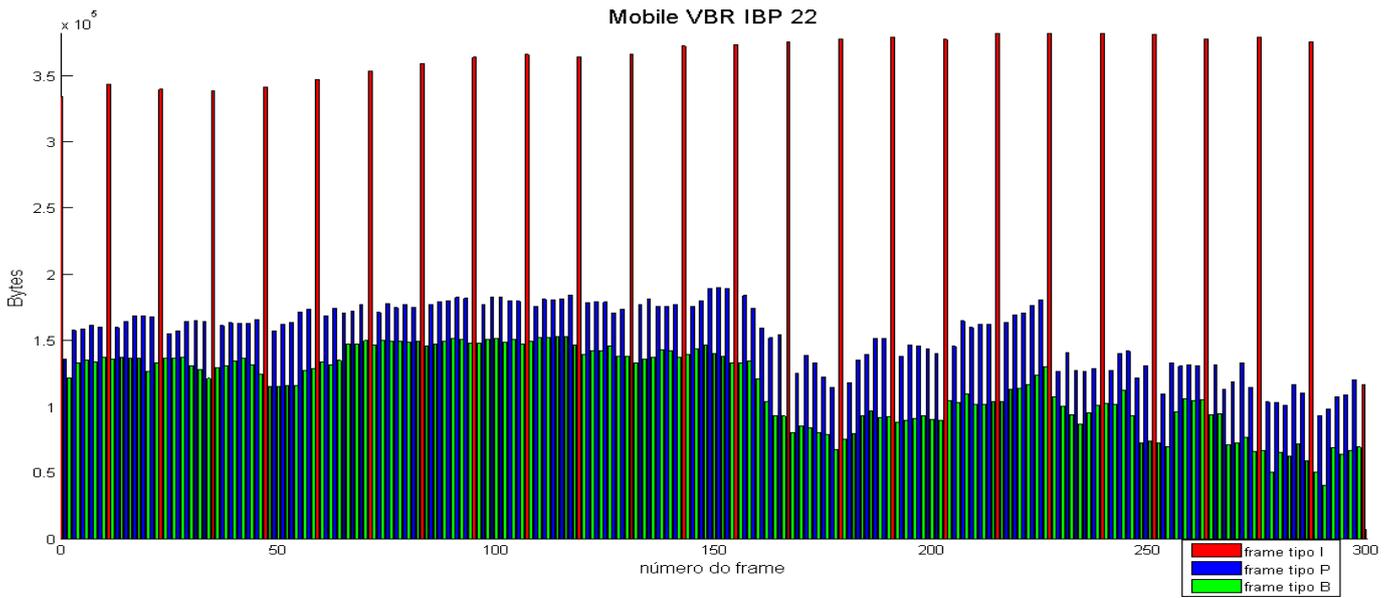


FIGURA 13 Frames I, B e P do vídeo Mobile com codificação VBR, GOP IBP e QP 22

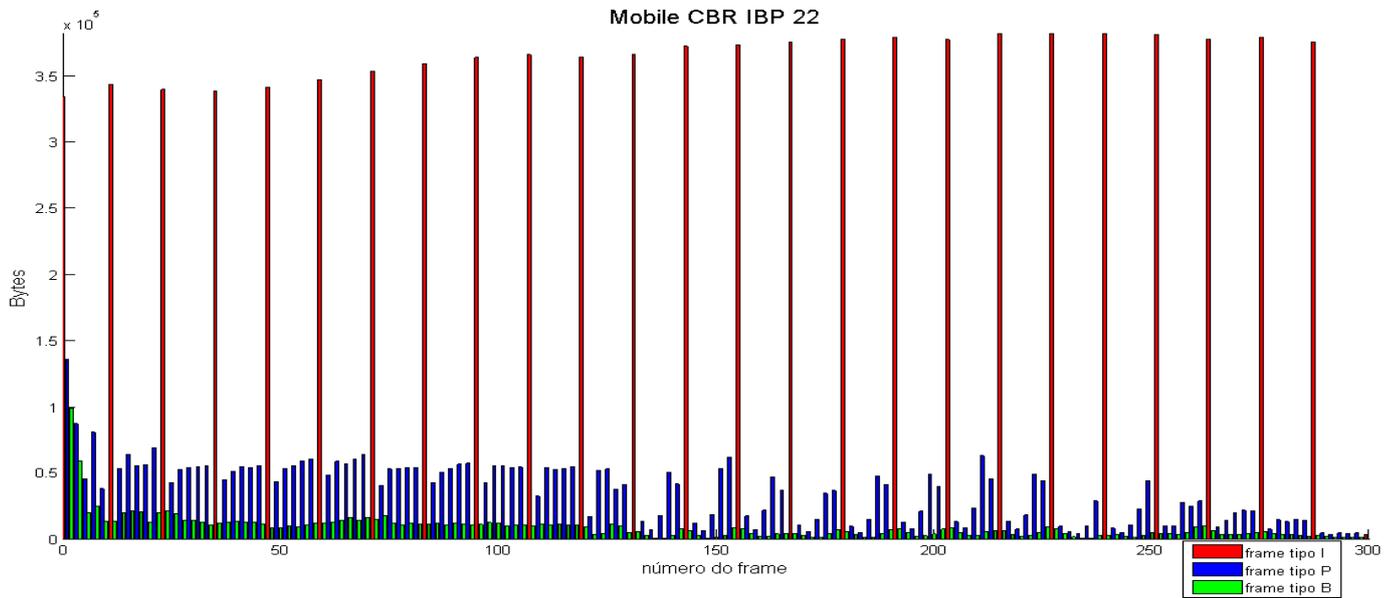


FIGURA 14 Frames I, B e P do vídeo Mobile com codificação CBR, GOP IBP e QP 22

A primeira característica importante nos gráficos, é a escala do eixo vertical, que representa o tamanho dos quadros em bytes. Enquanto os maiores quadros Akiyo, tipo I, são da ordem de 10.000 Bytes, os quadros Foreman são da ordem de 30.000 Bytes e os Mobile de 40.000 Bytes. Isto é um resultado direto da característica textura de cada vídeo influenciando na predição Intra. A textura mais homogênea do vídeo Akiyo, apresenta uma maior redundância espacial, enquanto no vídeo Mobile a textura varia bastante dentro de um quadro. Os quadros da sequência Foreman possuem características intermediárias.

A uniformidade do tamanho dos quadros tipo I da Akiyo se deve as características uniformes do vídeo em toda a sua extensão. A variação de tamanho dos quadros finais dos quadros Foreman se deve ao repentino redirecionamento da câmera. E a quase uniformidade do tamanho dos quadros Mobile deve-se à manutenção do nível de textura ao longo do vídeo.

Outra característica importante, e que era esperada, é que os frames do tipo P são em regra menores do que o tipo I, e do tipo B menores do que o tipo P. No gráfico Akiyo, os quadros P e B são muito reduzidos quando comparados aos quadros do tipo I. Isso também é um resultado da característica de baixa movimentação no vídeo, mas agora influenciando na predição Inter, pois os frames P e B necessitam de poucas informações novas para conseguirem ser codificados a partir dos seus frames de referência. O gráfico Mobile, por sua vez, apresenta muitos movimentos em praticamente todo o espaço de um quadro e em diversas direções. Logo, os seus quadros P e B tem tamanho de quase metade de um frame I. Finalmente o gráfico Foreman apresenta características intermediárias em relação aos outros.

Em resumo, a partir das características “bem comportadas” dos vídeos, como o Akiyo, conseguimos uma menor taxa e uma melhor PSNR, quando comparado com Foreman e principalmente Mobile, como podemos observar no gráfico de Taxa x Distorção para uma mesma configuração de codificação:

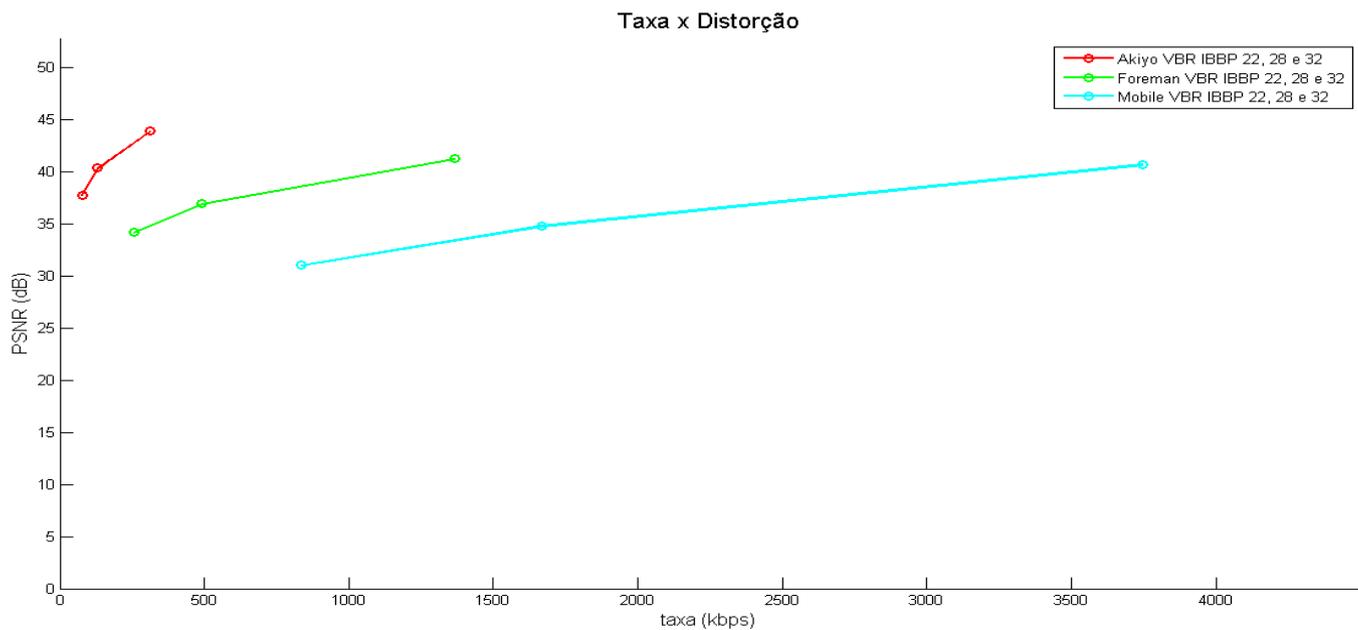


FIGURA 15 Gráfico Taxa x Distorção para os vídeos Akiyo, Foreman e Mobile com codificação VBR, GOP IBBP para os QP 22, 28 e 32, da esquerda para a direita

Após a codificação VBR dos vídeos e obtendo suas respectivas taxas médias, foram escolhidas as Taxas Alvo para a codificação CBR, onde o QP varia para que o codificador tente atingir a referida taxa. As Taxas Alvo para Akiyo e Foreman, foram escolhidas pouco abaixo das respectivas taxas VBR. Porém, as taxas VBR do Mobile eram altas demais para serem transmitidas em quaisquer das 4 redes utilizadas. Sendo assim, foram utilizadas as mesmas restrições escolhidas ao Foreman, o que reduziu bastante as taxas CBR Mobile em relação ao VBR, mas sob o risco de severa perda de qualidade do vídeo ainda no processo de codificação, o que pode-se observar na análise do PSNR dos frames apenas codificados:

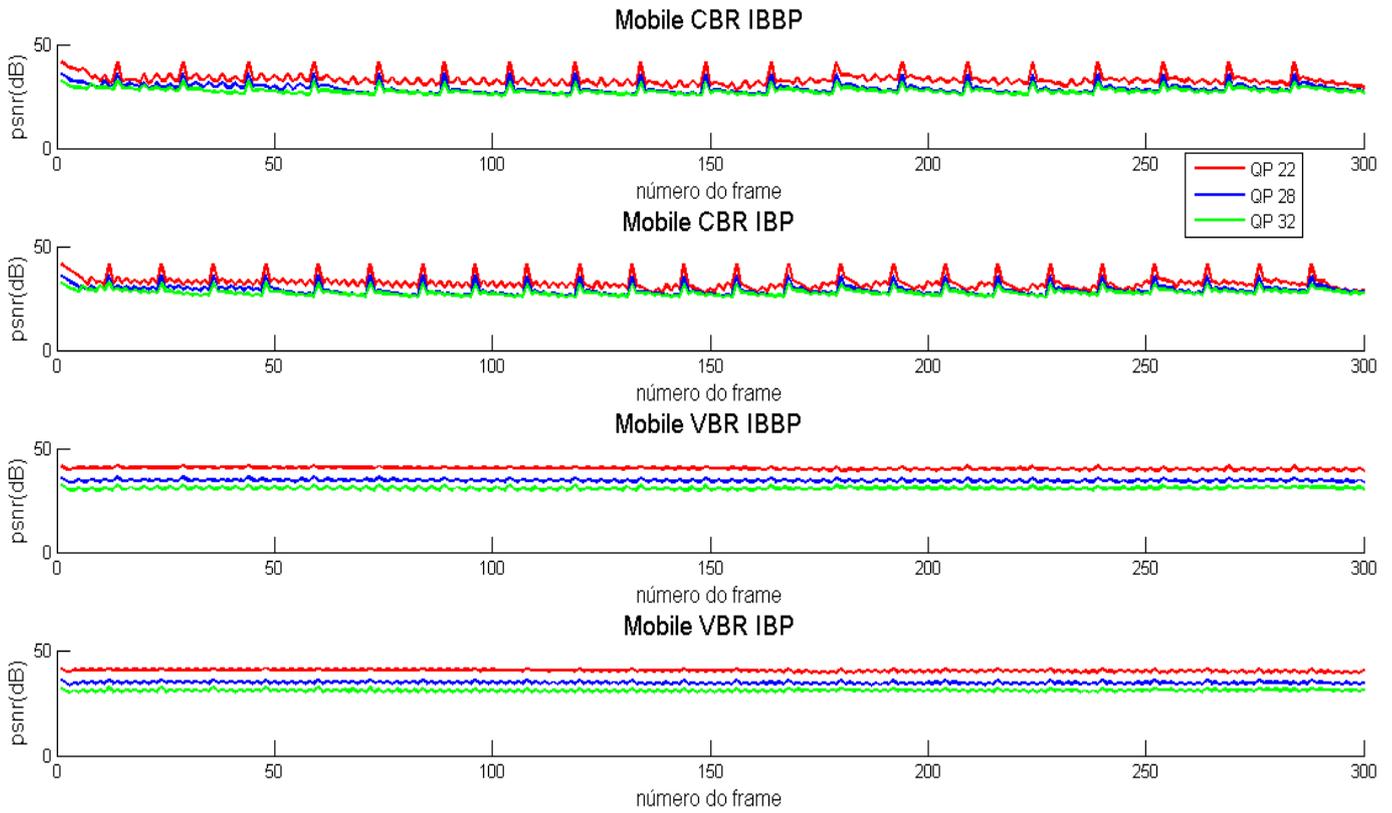


FIGURA 16 PSNR para diferentes codificações do vídeo Mobile

Com a codificação, os quadros foram divididos em pacotes de tamanho máximo de 1000 bytes. A quantidade média de pacotes em que foi dividido cada tipo de quadro para cada codificação está representado na TABELA 3:

		VBR						CBR					
		IBBP			IBP			IBBP			IBP		
		Akiyo	Foreman	Mobile									
I	QP 22	9,52	18,00	43,33	10,00	18,60	45,60	9,52	18,00	43,33	10,00	18,60	45,60
	QP 28	5,71	9,86	28,52	6,00	10,12	30,04	5,71	9,86	28,52	6,00	10,12	30,04
	QP 32	3,81	6,92	20,71	4,00	6,56	21,72	3,81	6,29	20,71	4,00	6,52	21,72
P	QP 22	2,12	8,91	19,42	1,80	7,73	19,57	2,61	8,76	6,03	2,20	7,70	4,82
	QP 28	1,07	3,67	9,47	1,00	3,11	9,45	3,13	3,36	1,57	2,67	2,97	1,40
	QP 32	1,00	2,00	5,15	1,00	1,78	5,03	1,04	1,85	1,18	2,74	1,67	1,15
B	QP 22	1,16	5,64	15,24	1,01	4,91	14,37	1,16	3,15	2,42	1,00	3,06	1,61
	QP 28	1,00	1,83	5,82	1,00	1,50	5,23	1,33	1,41	1,07	1,21	1,18	1,02
	QP 32	1,00	1,15	2,57	1,00	1,05	2,22	1,00	1,04	1,02	1,48	1,00	1,00

TABELA 3 Quantidade média de pacotes em que foi dividido cada tipo de quadro para os diferentes parâmetros de codificação

Vale a ressalva de que a representação de QP para codificações do tipo CBR remete ao parâmetro de quantização dos quadros tipo “I” e apenas este, pois nos demais, “P” e “B” o QP ajusta-se para atingir a taxa alvo do respectivo VBR de mesmo QP, porém este para todos os tipos de quadro.

As taxas atingidas pelas codificações Akiyo, Foreman e Mobile estão respectivamente expostos nas tabelas a seguir:

QP	GOP	VBR (kbit/s)	CBR (kbit/s)	Taxa alvo (kbits/s)
22	IBBP	311.28	327.84	300
22	IBP	326.32	336.10	300
28	IBBP	130.52	135.67	120
28	IBP	144.66	137.44	120
32	IBBP	77.25	79.89	70
32	IBP	88.01	81.36	70

TABELA 4 Taxas médias atingidas pelos vídeos Akiyo codificados

Observa-se que mesmo ao ser somado o tráfego de concorrente de 518,4 kbps no pior caso, ou seja, espera-se que todas as codificações consigam ser transmitidas sem perdas inclusive na rede 3 de maior restrição (de 0.9Mbps/1Mbps e tamanho de *buffer* de 40 pacotes).

QP	GOP	VBR (kbit/s)	CBR (kbit/s)	Taxa alvo (kbits/s)
22	IBBP	1368.20	1158.37	1100
22	IBP	1351.62	1185.00	1100
28	IBBP	489.25	472.29	400
28	IBP	490.52	442.40	400
32	IBBP	256.04	219.50	200
32	IBP	257.05	228.43	200

TABELA 5 Taxas médias atingidas pelos vídeos Foreman codificados

No vídeo Foreman, com o QP de 22 ultrapassamos a banda máxima ainda na saída do transmissor o que nos dá certeza de perdas de pacotes. Com o QP 28 consegue-se transpor as redes 1 e 2 de maior banda, porém nas redes 3 e 4 a taxa está muito próximo ao banda passante máxima da rede. Neste caso a análise se torna delicada, pois o tamanho de *buffer* da rede 4 em relação a rede 3 é relevante, como se observa nos gráficos de perda a seguir, donde observamos que as perdas mais severas acontecem na codificação VBR, sem controle de taxa, e está relacionado ao instante em que a câmera muda de direção. As posições dos pacotes perdidos estão representadas pelas linhas vermelhas nas FIGURAS 17 a 22:

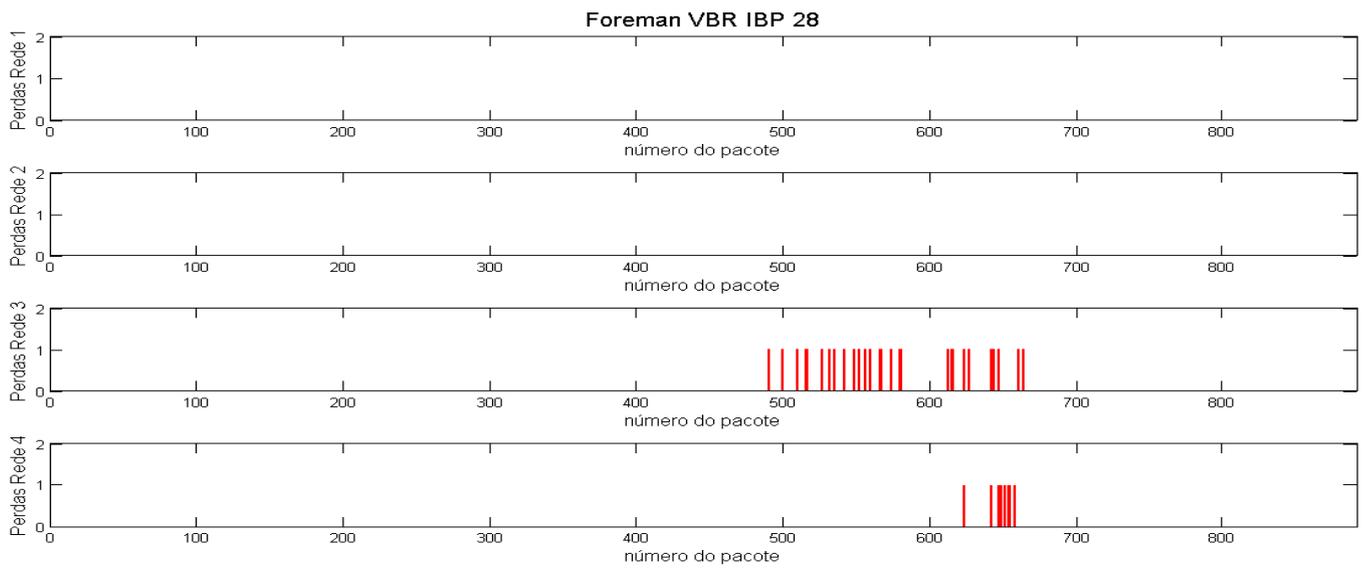


FIGURA 17 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Foreman com codificação VBR, GOP IBP e QP 28

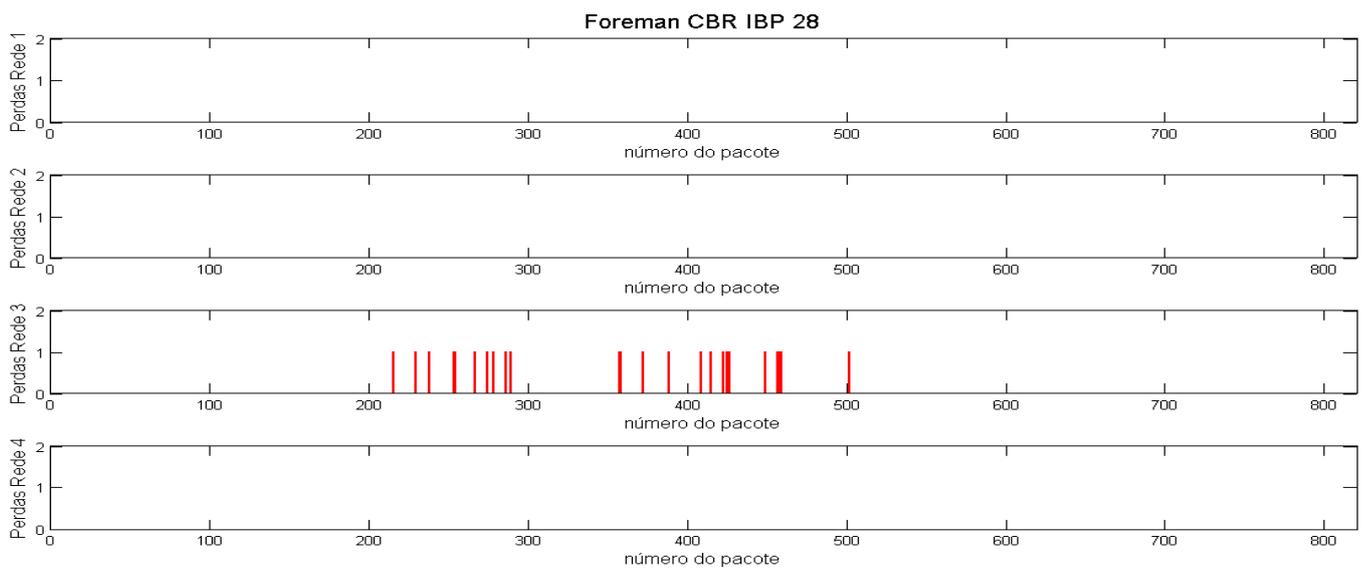


FIGURA 18 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Foreman com codificação CBR, GOP IBP e QP 28

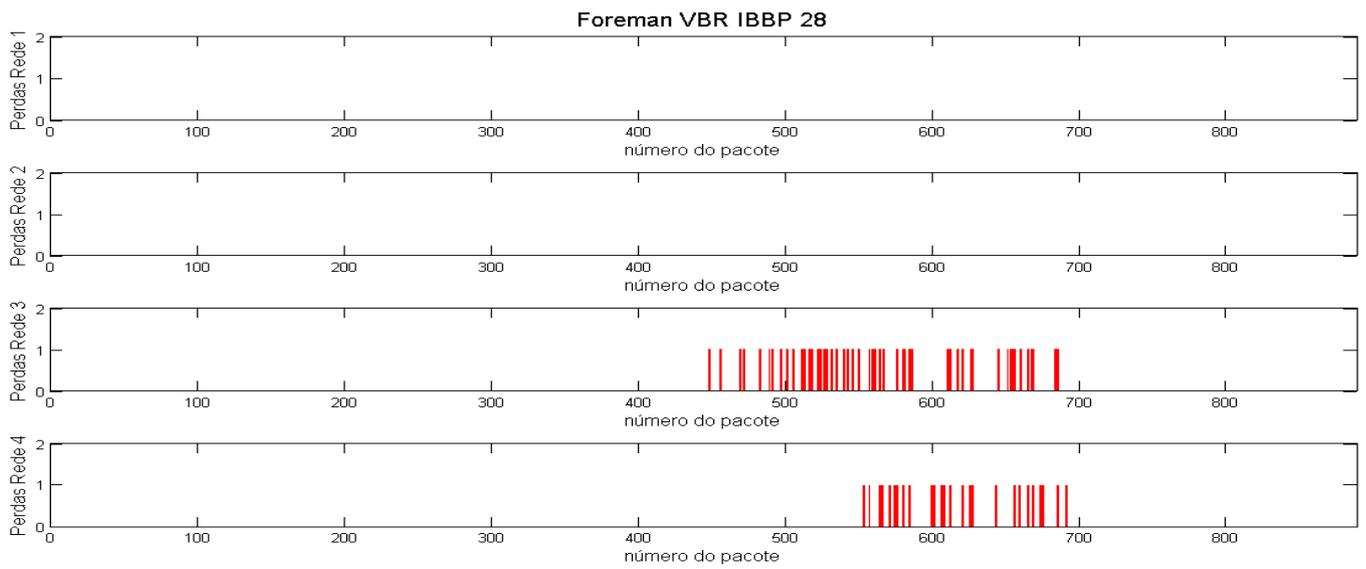


FIGURA 19 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Foreman com codificação VBR, GOP IBBP e QP 28

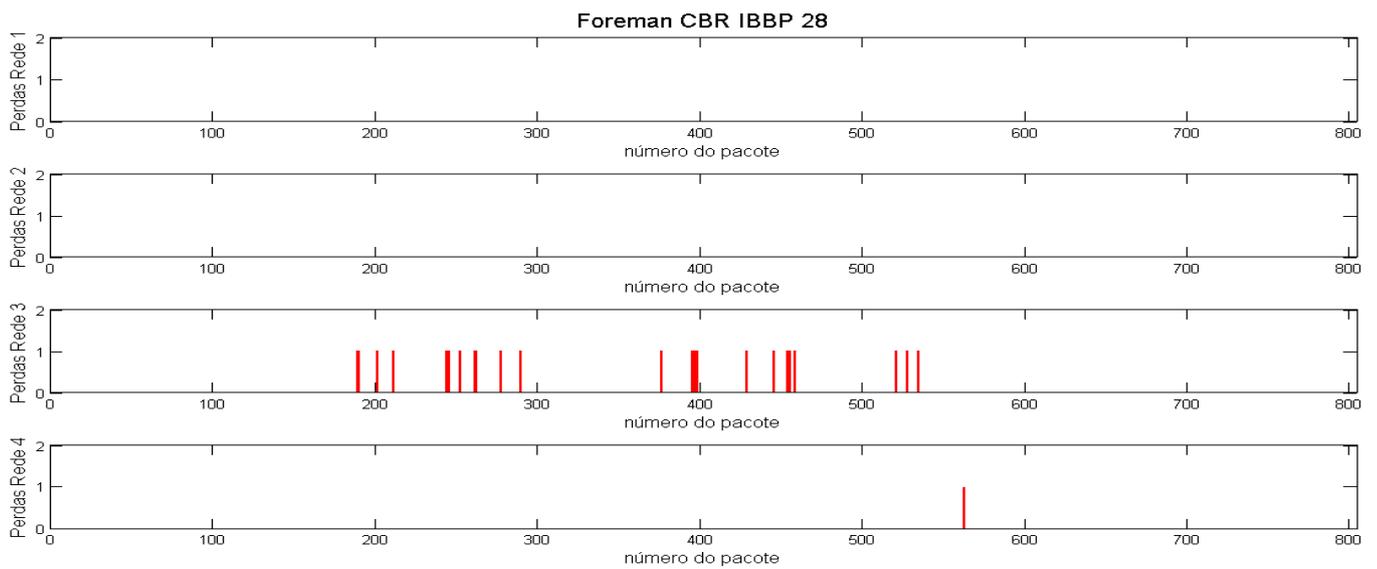


FIGURA 20 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Foreman com codificação VBR, GOP IBP e QP 28

Com o QP de 32 a análise é semelhante a feita no caso Akiyo, ou seja, conseguem ser transmitidos em quaisquer redes.

QP	GOP	VBR (kbit/s)	CBR (kbit/s)	Taxa alvo (kbits/s)
22	IBBP	3747.77	1220.55	1100
22	IBP	3808.65	1246.63	1100
28	IBBP	1669.21	527.70	400
28	IBP	1747.27	607.88	400
32	IBBP	835.75	373.10	200
32	IBP	914.51	441.20	200

TABELA 6 Taxas médias atingidas pelos vídeos Mobile codificados

Nota-se que todas as codificações Mobile VBR sofreram perdas ainda na saída do transmissor, bem como os CBR com QP de 22, pois, neste caso o grande tamanho dos frames I fragmentados na ordem de 50 pacotes, saturariam imediatamente o *buffer* de tamanho 40, e não esvaziaria a antes da chegada de outros frames, tipo P da ordem de 20 e tipo B da ordem 15 pacotes, saturando em poucos instantes o *buffer* de transmissão. A TABELA 7 nos ilustra que em único quadro Mobile, no caso logo o primeiro, pode extrapolar a capacidade do *buffer*. Além disto, a análise de QP de 28 recai novamente na análise do tamanho de *buffer*, porém o principal problema se dá no nó 2 de convergência dos fluxos:

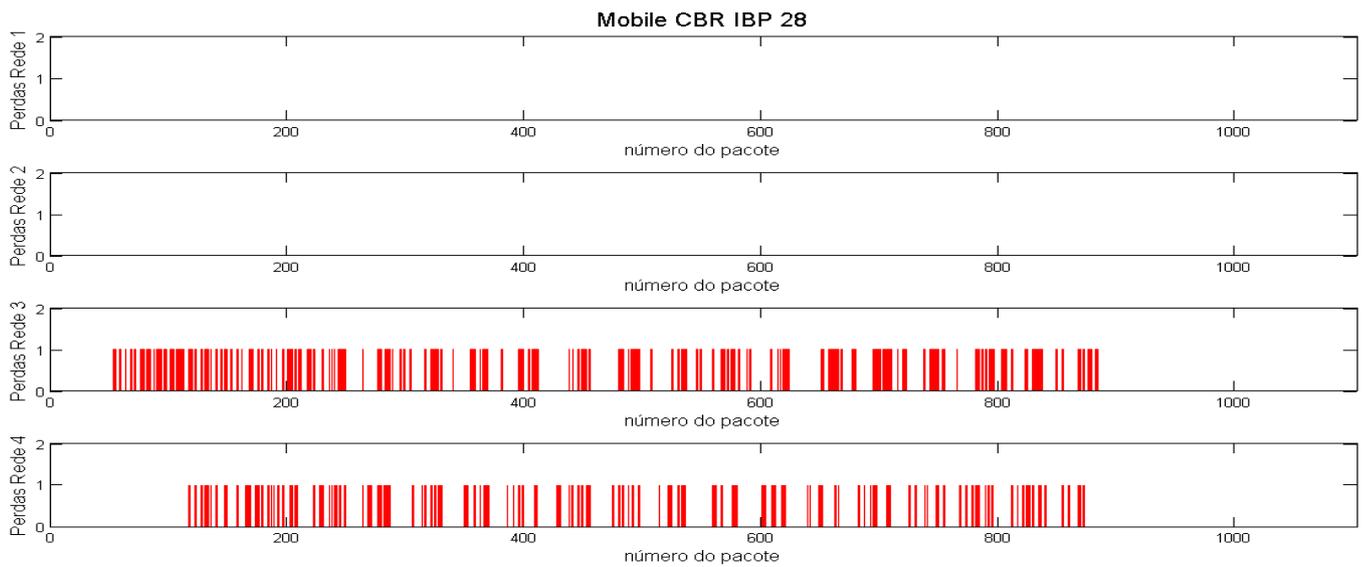


FIGURA 21 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Mobile com codificação CBR, GOP IBP e QP 28

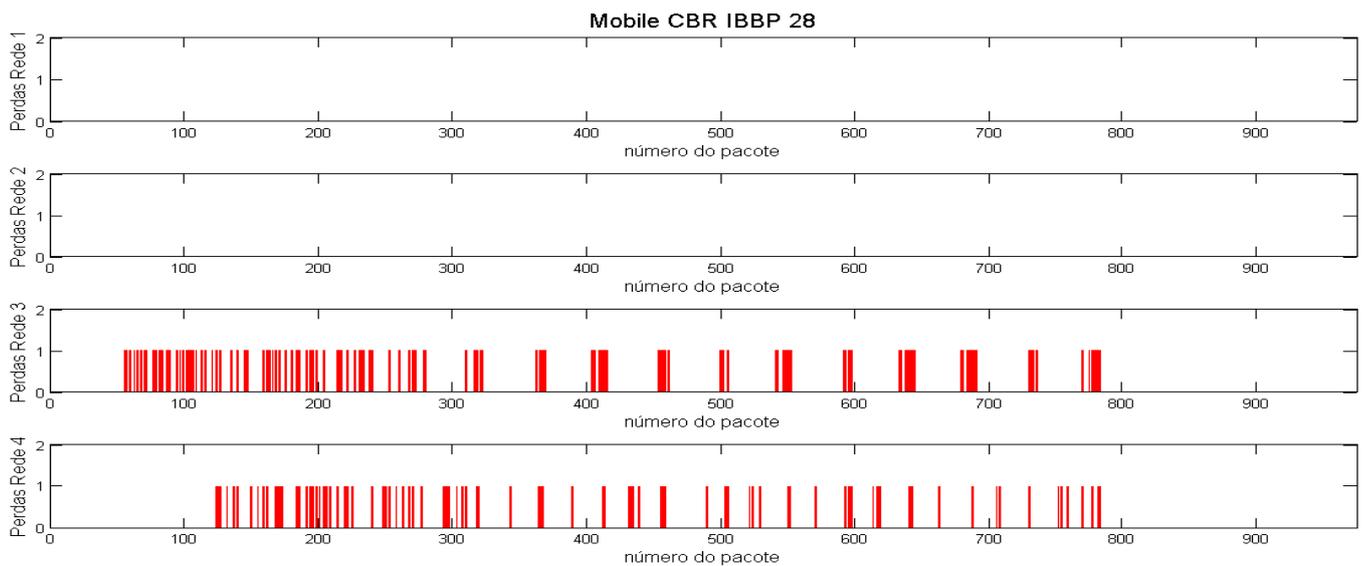


FIGURA 22 Pacotes perdidos nas redes 1, 2, 3 e 4 do vídeo Mobile com codificação CBR, GOP IBBP e QP 28

Porém nota-se certa periodicidade nas perdas, que se devem justamente ao grande tamanho do frame I, que em um único instante chega a mandar para o *buffer* de transmissão da ordem de 30 pacotes, que ao chegarem ao *buffer* de compartilhamento de

tráfego, concorrerão com os pacotes do fluxo constante ocasionando a saturação das *buffer* e perdas de pacotes.

Com base nestes resultados preliminares, podemos fazer comparar a perda de pacotes com o gráfico Taxa x distorção para as diferentes codificações, bem como a análise do PSNR ao decorrer dos frames. O exemplo a seguir ilustra a codificação do Foreman CBR de GOP IBBP e QP de 22. Os valores de PSNR iguais a 55 representam ao pacotes não perdidos, ou seja, saíram da rede exatamente como entraram. As perdas de pacote são implicam em queda de qualidade, ou seja, dos valores do PSNR. Observa-se no gráfico as quedas mais recorrentes na rede 3, pior caso, e uma pequena queda de qualidade na rede 4, concordando com o respectivo gráfico de perda de pacotes.

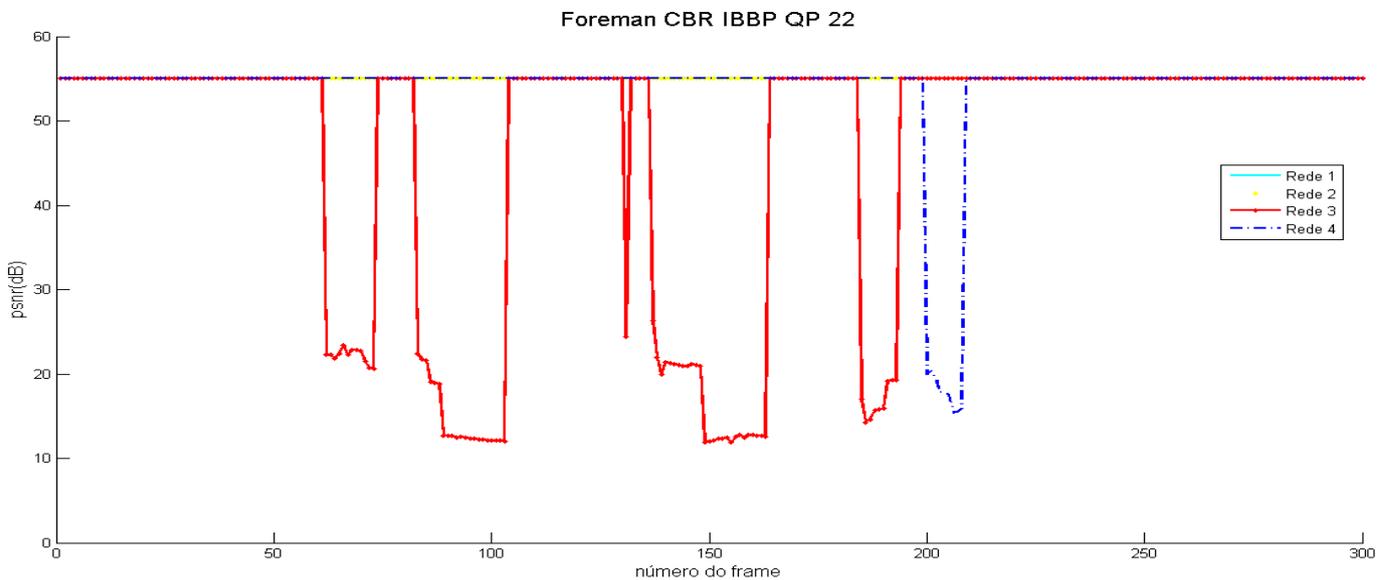


FIGURA 23 Queda de PSNR devido a perdas de pacotes nas redes 3 e 4 do vídeo Foreman com codificação CBR, GOP IBBP e QP 22

7- CONCLUSÃO

A compressão de vídeo depende fortemente das características do vídeo, textura e movimento. Quanto mais informação puder ser reutilizada, ou seja, retirada as redundâncias, mais eficiente será a compressão. Isto pôde ser observado quanto se compara o tamanho dos frames dos vídeos Akiyo, Foreman e Mobile, assim em ordem crescente. Também pode ser observado na FIGURA 15, em que para o mesmo PSNR a taxa das codificações Akiyo chegam a ser 10 vezes menores do que as taxas da codificação Mobile, para a mesma codificação e tamanho de quadros I até 3000 vezes menores.

Este aumento de taxa pode problemático quanto a largura de banda de rede que deseja-se transmitir o vídeo. Para a redução desta taxa, pode-se ativar o controle de fluxo, ou seja, buscar características de Taxa de Bits Constante, e desta forma limitar o fluxo de saída de dados. As principais perdas de PSNR são percebidas nos quadros P e B, pois os quadros tipo “I” foram tiveram sua o parâmetro de quantização pré-determinado. O uso dessa técnica com fortes restrições de taxa pode haver fortes reduções perceptíveis tanto quantitativamente, pela redução de PSNR como mostrado na FIGURA 16, quanto visualmente.

Finalmente, ainda que a taxa média de do vídeo seja menor do que o limite da banda, o fluxo de bits não é constante na saída do transmissor. Os quadros do tipo “I”, geralmente muito maiores do que os demais, quando são empacotados e os pacotes armazenados no *buffer* para a transmissão a uma taxa suportada pela rede, podem satura-lo e havendo perdas de pacote já na transmissão. Outras perdas sofridas podem ser causadas por congestionamento em outros pontos da rede. Desta forma, a codificação deve estar também em conformidade com o dimensionamento dos *buffers*.

8- REFERÊNCIAS BIBLIOGRÁFICAS

1. ISO/IEC 14496-10:2012 Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding
2. J. KLAUE; B. RATHKE; A. WOLISZ; EvalVid – A Framework for Video Transmission and Quality Evaluation. 13th International Conference on Modeling Techniques and Tools for Computer Evaluation. Urbana, Illinois, USA, 2003.
3. A. TANENBAUM. Computer Networks. 4th Edition. Prentice Hall, 2003.
4. RICHARDSON; IAIN. The H.264 advanced video compression standard. 2 ed. United Kingdom: Wiley editor, 2010.
5. TORRES, Daniel Pioli. Impacto do Gerenciamento Ativo de Filas em Redes IP Transportando Fluxo de Vídeo h.264/MPEG-4 AVC. Curitiba, 2009.
6. NS-2 Network Simulator, <http://www.isi.edu/nsnam/ns/>
7. EvalVid - A Video Quality Evaluation Tool-set, <http://www.tkn.tu-berlin.de/research/evalvid/>
8. H.264/AVC JM Reference Software, <http://iphome.hhi.de/suehring/tml/>