

# Diffusion de Vidéo hiérarchique sur Internet : Évaluation du codeur PVH

- David PATE, pate@dpt-info.u-strasbg.fr  
Université Louis Pasteur, Illkirch
- Stella MARC-ZWECKER, marc-zwecker@dpt-info.u-strasbg.fr  
Université Louis Pasteur, Illkirch
- Jean-Jacques PANSIOT, pansiot@dpt.info.u-strasbg.fr  
Université Louis Pasteur, Illkirch

*Le codage hiérarchique de vidéo lors d'une diffusion permet de s'adapter aux besoins ou capacités des récepteurs et du réseau. Nous présentons le principe du codage hiérarchique et plus particulièrement l'algorithme PVH (Progressive Video with Hybrid transform). Nous offrons une évaluation de ce système de codage.*

## ■ Introduction

Ces dernières années, la diffusion sur Internet s'est considérablement déployée par le biais de IP (Internet Protocol) multipoint. En se greffant sur l'Internet point-à-point, il s'est créé un véritable réseau virtuel, appelé Multicast Backbone (ou MBone), permettant la diffusion.

Les besoins en qualité lors d'une diffusion vidéo sont très divers et dépendent de plusieurs facteurs. On trouve parmi ceux-ci le niveau de qualité désiré par les récepteurs, mais aussi des contraintes matérielles tels que les puissances de codage/décodage des stations émettrices/réceptrices, le débit et la qualité de service disponibles tout au long du chemin de diffusion.

La diffusion hiérarchique permet de s'adapter aux besoins de chaque destinataire. Ainsi, chaque flux est décomposé en sous-flux complémentaires. Chaque sous-flux apporte un peu plus d'information. Chaque empilement de flux correspond à un niveau de qualité.

La partie 2 présente les techniques de codage de la vidéo. La partie 3 explique les mécanismes du codage PVH<sup>1</sup> (Progressive Video with Hybrid transform) qui est un codeur permettant le codage hiérarchique, la partie 4 explique les principes de la diffusion et la partie 5 offre une évaluation des flux générés par PVH.

## ■ Le codage vidéo

### Outils de codage vidéo

Dans ce chapitre, nous présentons quelques principes utilisés dans le codage vidéo. Premièrement, la plupart des algorithmes utilisent la décomposition en luminance (intensité lumineuse) et chrominance (couleur) du flux vidéo. Cela correspond à un changement de base par rapport au codage en rouge-vert-bleu. Les transformations s'appliquent principalement à la luminance dont les différences sont les mieux perçues par l'œil humain.

#### **DCT**

La transformée en cosinus discrète (DCT : Discrete Cosine Transform) a pour but de décorrélérer les coefficients d'une image. En effet, les pixels voisins sont souvent liés, car ils représentent un même objet. Or, la DCT va rompre cette liaison pour regrouper les informations de fréquences de l'image. Dans l'équation 1,  $f(i,j)$  est le coefficient avant transformation, et  $F(u,v)$  celui après transformation.

La composante continue  $F(0,0)$  est celle qui a le plus de valeur. En effet, la DCT concentre les coefficients de magnitude élevée dans la partie supérieure gauche de la matrice. Les autres valeurs sont les composantes alternatives.

---

<sup>1</sup> PVH est inclus dans *mash* (<http://www-mash.cs.berkeley.edu/mash/index.html>).



$$F[u, v] = \frac{1}{4} \sum_{i,j} \lambda(u) \lambda(v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} f(i, j)$$

$$\lambda(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{pour } \xi = 0 \\ 1 & \text{pour } \xi \neq 0 \end{cases}$$

Equation 1. Formules de la DCT.

### La quantification

La quantification a pour but de diminuer la longueur des messages. Elle consiste en une troncature des champs de bits. Cela équivaut à diviser un coefficient par une puissance de deux. Cela entraîne une perte d'information.

La quantification uniforme consiste à diviser tous les coefficients par la même valeur. Mais il est judicieux d'adapter le quantificateur. On crée donc des tables de quantifications qui enlèvent peu d'informations aux coefficients importants (cf. DCT).

### La décomposition en ondelettes

La décomposition en ondelettes est très proche de la transformée en cosinus discrète. On applique récursivement des filtres passe-haut ( $H$ ) et passe-bas ( $L$ ) aux images, comme on peut le voir dans les figures 1 et 2. On obtient ainsi une image multi-résolutions.

$$L(i) = \lfloor (c[2i] + c[2i+1]) / 2 \rfloor$$

$$H(i) = c[2i] - c[2i+1]$$

Equation 2 : exemple de filtres passe-haut et passe-bas

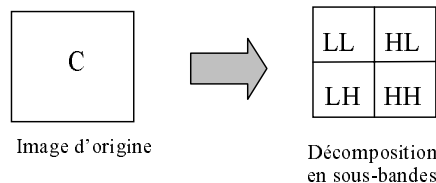


Figure 1. Décomposition en ondelettes.

Ce type de codage permet aisément de créer un flux hiérarchique puisque chaque niveau de transformation peut être considéré comme un niveau de la hiérarchie. flux spatiaux, on commence par déterminer les blocs à coder par reconstruction conditionnelle comme dans intra-H261. Puis, on opère une décomposition en ondelettes des blocs ainsi sélectionnés. La bande LL qui est une application récursive de deux fois le filtre passe-bas, est décomposée par un codage DCT en couche (Layered DCT[1]). Les bandes LH et HL sont codées en sous-bandes par un arbre quaternaire (Layered SBC (SubBand Coding)). HH est abandonnée, car les informations qu'elle contient n'ont que peu de valeur lors du décodage.

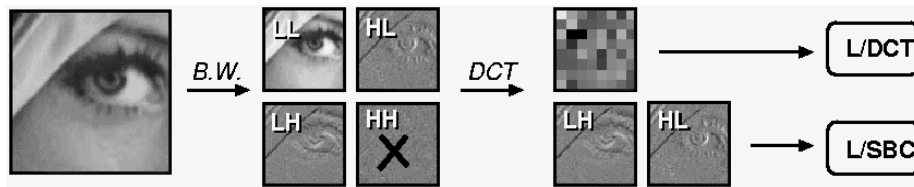


Figure 2. Décomposition d'une image par PVH.

Pour créer des sous-flux, PVH utilise des passes de raffinement. Ainsi, dans le cas de la DCT hiérarchique, le flux de base contient la composante continue plus les composantes alternatives quantifiées de chaque bloc à coder. Les flux complémentaires sont obtenus en émettant les composantes alternatives avec un quantificateur moins important. Ceci revient à envoyer un bit de plus d'information par composante.

Une fois que les données vidéo ont été codées, il faut les mettre dans des paquets pour les envoyer sur les réseaux informatiques. Pour cela, on utilise les techniques décrites dans la partie « La diffusion hiérarchique ». On se sert plus particulièrement de RTP qui permet de diffuser des flux temps-réel comme l'audio et la vidéo. Les différentes couches de codage sont réparties dans différents groupes multipoints. On peut, par exemple, décider de regrouper dans un même groupe les couches de base de la luminance et celle de la chrominance ou au contraire de les mettre dans des groupes différents. De ce fait, dans le premier cas, en s'abonnant à un unique groupe, on réceptionne de la vidéo en couleur. Dans le deuxième cas, il faut s'abonner à deux groupes pour recevoir les mêmes flux. Il en est de même pour tous les flux de raffinement qui peuvent être regroupés ou non.

En fonction de la répartition des flux vidéo dans les groupes réseau, il faut s'abonner à plus ou moins de groupes pour recevoir la même qualité d'image. Si l'on décide de disperser les flux dans de nombreux groupes, les récepteurs ayant peu de ressources que ce soit réseaux ou calculateurs ont la possibilité de s'abonner à moins de groupes. De ce fait, ils décodent une vidéo de moindre qualité, alors qu'ils n'auraient rien réceptionné si tous les flux étaient arrivés en même temps dans le même groupe.

Nous venons de voir la technique de codage vidéo PVH et plus particulièrement sa capacité à générer un flux hiérarchique. Nous allons voir maintenant les techniques pour diffuser ces flux hiérarchiques sur Internet.

## ■ La diffusion hiérarchique

### Les principes de la diffusion

Lorsqu'une source émet un paquet en direction d'un groupe, celui-ci va parcourir un arbre de diffusion. Ce dernier peut être différent pour chaque source comme dans DVMRP[4], PIM-DM[5], MOSPF[6] ou au contraire commun à toutes les sources à la manière de CBT[7], PIM-SM[8]. Les membres d'un groupe n'ont généralement pas connaissance de la topologie de l'arbre.

Du point de vue de la source, l'émission d'un paquet multicast est identique à celle d'un paquet point-à-point. Ce sont les routeurs multipoint qui se chargent de la diffusion des paquets. Les feuilles de l'arbre sont alors les récepteurs<sup>2</sup>.

Une application vidéo est assez robuste pour subir des pertes et/ou des déséquences. Cela est dû à la nature même de la vidéo. En revanche, elle requiert des délais faibles. C'est pourquoi, on utilise généralement des transmissions sans connexion. Les contrôles sont davantage effectués au niveau de l'application plutôt qu'au niveau réseau, car seules les couches supérieures (au sens OSI) prennent en compte ces propriétés.

### RTP/RTCP

Pour effectuer une diffusion de paquets provenant d'une source temps réel, on utilise de plus en plus RTP et RTCP[9]. Les deux protocoles sont généralement gérés par l'application.

RTP est un protocole de bout en bout utilisé par les sources, qui permet l'identification du type de données, la numérotation des séquences et l'estampillage des paquets. Il ne garantit en rien l'ordre ni l'heure de livraison.

Le suivi de la livraison est assuré par RTCP. Les récepteurs fournissent des rapports à la source par son intermédiaire. Pour limiter le trafic de contrôle, RTCP ne s'autorise l'utilisation que d'un certain pourcentage de la bande passante. Donc, plus le groupe grandit, moins les rapports d'un récepteur seront fréquents.

<sup>2</sup> Un nœud intermédiaire peut également être récepteur.



## La diffusion hiérarchique

Pour permettre de satisfaire le plus grand nombre de récepteurs dont les capacités sont diverses, on se sert de la diffusion à qualité variable. Pour effectuer ce type de diffusion, nous utilisons le codage PVH. La source a la possibilité d'émettre un flux hiérarchique (un flux de base et un flux complémentaire<sup>3</sup>) pour s'adapter aux ressources des récepteurs. Cela signifie que le flux d'origine est décomposé en sous-flux ce qui permet de le reconstruire en les agrégeant. Une partie seulement de ces sous-flux peut servir à recomposer un signal de moins bonne qualité que celui d'origine.

Les récepteurs ne voulant ou ne pouvant pas avoir la meilleure qualité ne demandent qu'à recevoir le flux de base. La source émet dans tous les cas l'ensemble des flux. Ce sont les nœuds intermédiaires, en fonction de la demande des sous-arbres qu'ils gèrent, qui laissent passer ou non un flux. De ce fait, un flux particulier n'est pas rediffusé dans les sous-arbres si personne ne le réclame pour économiser des ressources.

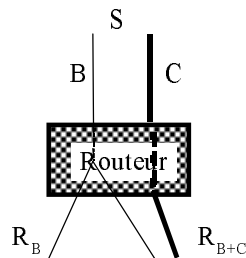


Figure 3. Arbre de diffusion.

Sur l'exemple de la figure 3, nous considérons une source S qui émet un flux décomposé en deux sous-flux, celui de base (B) et le complémentaire (C). Nous avons deux récepteurs  $R_B$  et  $R_{B+C}$ , l'un demandant le flux de base (B), et l'autre voulant l'intégralité du signal (B+C). La source émet chaque sous-flux dans un groupe différent. De ce fait, pour obtenir le flux complet, il faut être abonné aux deux groupes. Ce mécanisme élimine la redondance. En revanche, elle peut engendrer des déséquilibrages puisque les flux empruntent des chemins différents. Il faut donc, en permanence gérer la synchronisation entre les sous-flux.

Après avoir étudié les mécanismes qui permettent de diffuser les flux hiérarchiques sur Internet, voyons les caractéristiques des flux générés par le codeur PVH.

## ■ Evaluation des flux PVH

### Modalités des tests

Pour réaliser nos tests, nous avons filmé en format QCIF (176x144 pixels) soit un écran noir pour un mouvement minimal, soit un orateur, censé engendrer un trafic moyen, soit un ventilateur dont le mouvement met à jour toute l'image plusieurs fois par seconde pour un trafic quasiment maximal, soit la neige d'un syntoniseur de télévision pour un trafic encore plus important puisque chaque pixel de l'image est continuellement renouvelé. Nous avons émis nos données dans 16 groupes multipoint différents dont les identifiants sont composés d'une adresse IP et d'un numéro de port dont les valeurs s'échelonnent de 50 000 à 50 031. Les numéros pairs de port contiennent des données RTP, et les numéros impairs servent aux flux de contrôle RTCP. Suivant les tests, tous les sous-groupes ne sont pas utilisés. L'affectation des numéros de port aux sous-groupes des tests dont on trouve les résultats dans les figures 7 et 8, sont résumés dans le tableau 1.

Tableau 1. Description des flux utilisés dans les tests (fb : flux de base ; fr : flux de raffinement).

<sup>3</sup> Le flux complémentaire peut lui-même être découpé en sous-flux. Chaque sous-flux identifie un niveau de qualité

Numéros de ports	Un sous-flux pour chaque raffinement	Un sous-flux pour tous les raffinements
50000	fb de la luminance (en DCT)	fb de la luminance (en DCT)
50002	fr de la luminance de 1 bit	fr de la luminance de 3 bits
50004	fr de la luminance de 1 bit	
50006	fr de la luminance de 1 bit	
50008	fb de la chrominance (en DCT)	fb de la chrominance (en DCT)
50010	fr de la chrominance de 1 bit	fr de la chrominance de 3 bits
50012	fr de la chrominance de 1 bit	
50014	fr de la chrominance de 1 bit	
50016	fb du codage en sous-bande	fb du codage en sous-bande
50018	fr de 2 passes du codage en sous-bande	fr de 4 passes du codage en sous-bande
50020	fr de 2 passes du codage en sous-bande	

### Caractéristiques des flux PVH

Première remarque : si l'on filme en format CIF (352x288 pixels), la station<sup>4</sup> qui sert à coder la vidéo ne permet pas de dépasser une douzaine d'images par secondes. En QCIF, on atteint 30 images/s.

La figure 4 nous montre que le débit RTCP reste constant. Au regard des autres résultats, on se rend compte que le trafic des données n'influe pas sur le contrôle dont le débit global est aux alentours de 5Kb/s, ce qui correspond aux spécifications de RTCP.

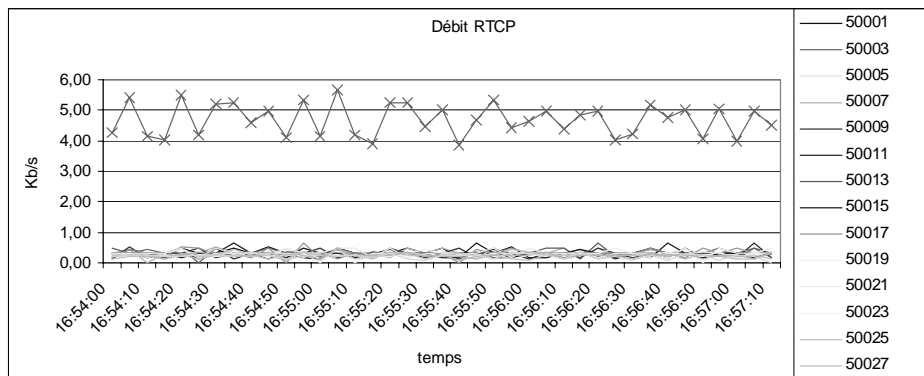


Figure 4. Débit RTCP.

Les figures 5a et 5b permettent de voir la différence de qualité entre une image avec et sans les flux de raffinement. Sur la figure 5a, on peut distinguer l'objet filmé, mais seule la figure 5b autorise un suivi prolongé de la vidéo.



Figure 5. (a) Image obtenue avec uniquement le flux de base de la luminance  
 (b) image obtenue avec le flux de base + un raffinement de 3 bits de la luminance.

A l'œil nu, il n'est pas possible de faire une différence de qualité sur les images avec ou sans les flux du codage en sous-bande. Seules les images obtenues en récupérant la neige d'un syntoniseur de télévision non réglé montrent une différence de qualité entre des images avec et sans codage en sous-bande. Or sans codage en

<sup>4</sup> PC sous FreeBSD muni d'un AMD K6-II à 350 MHz avec 64Mo de mémoire vive.



sous-bande, le trafic est d'environ 1 Mb/s alors qu'il atteint 3,5 Mb/s avec les flux du codage en sous-bande. C'est pourquoi, dans les tests suivants nous ne parlerons plus de ce type de codage qui engendre un flux très important pour un apport de qualité minime.

Afin d'évaluer le codage PVH, nous le comparons au codage intra-H261, très utilisé dans la diffusion de vidéoconférence. A l'étude des figures 6a, 6b, 6c et 6d, on se rend compte que plus la vidéo est animée, plus les débits engendrés par PVH sont moindres que ceux provenant du codeur intra-H261. Nous avons pris un quantificateur de valeur 1 pour intra-H261 et tous les flux de raffinement possibles (sauf le codage en sous-bande) pour PVH pour obtenir la meilleure qualité d'image possible à un taux de rafraîchissement de 30 images par seconde. La qualité visuelle est quasiment identique avec un très léger avantage pour intra-H261. Mais, en aucun cas, la différence de qualité ne justifie de tels écarts de débit.

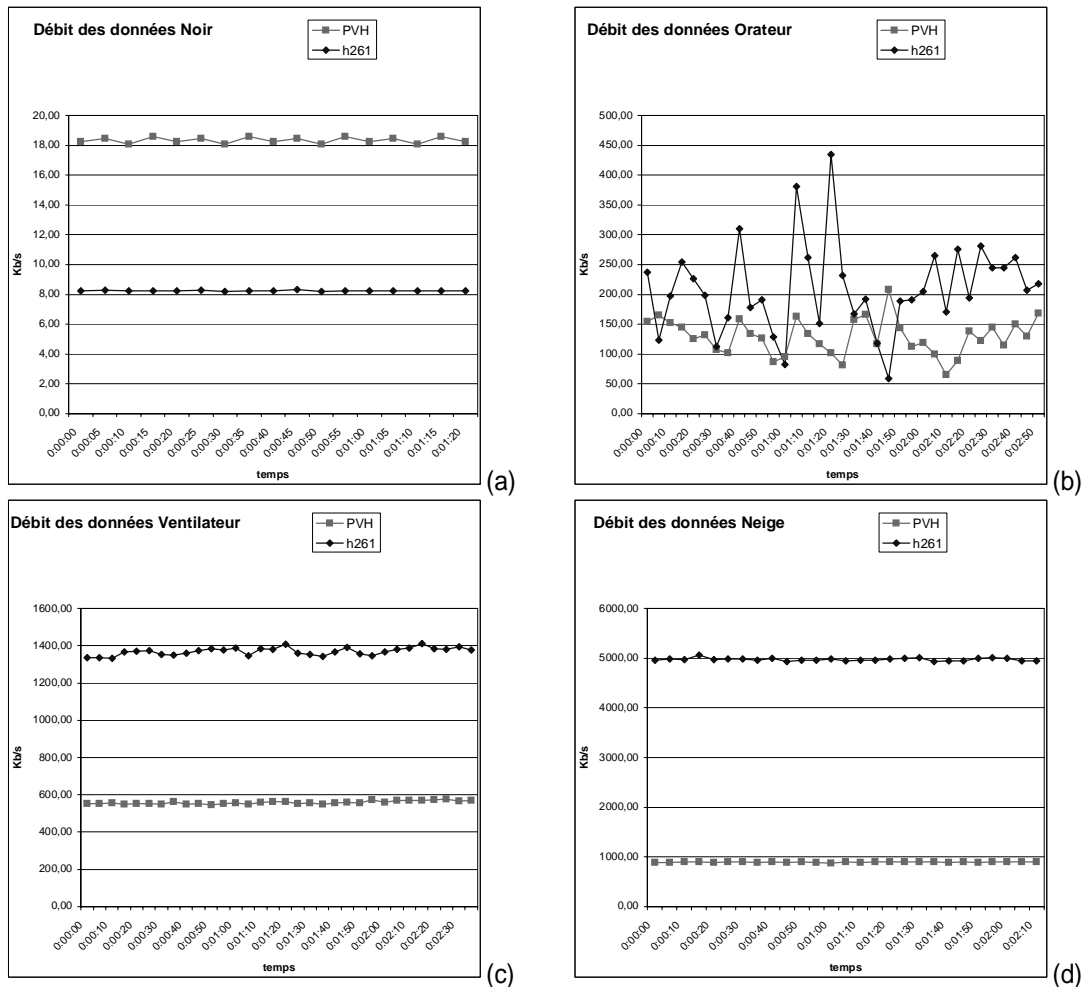


Figure 6. Comparaison des débits engendrés par PVH et intra-H261 en filmant (a) du noir, (b) un orateur, (c) un ventilateur et (d) un syntoniseur télé.

La figure 7 montre que si l'on cumule les flux de luminance d'une vidéo moyennement animée, on atteint environ 80 Kb/s. Cette valeur est très correcte pour faire de la vidéoconférence sur le Mbone. De plus, le débit moyen du flux de base est d'environ 25 Kb/s ce qui devrait permettre de recevoir les données par l'intermédiaire d'un modem standard.

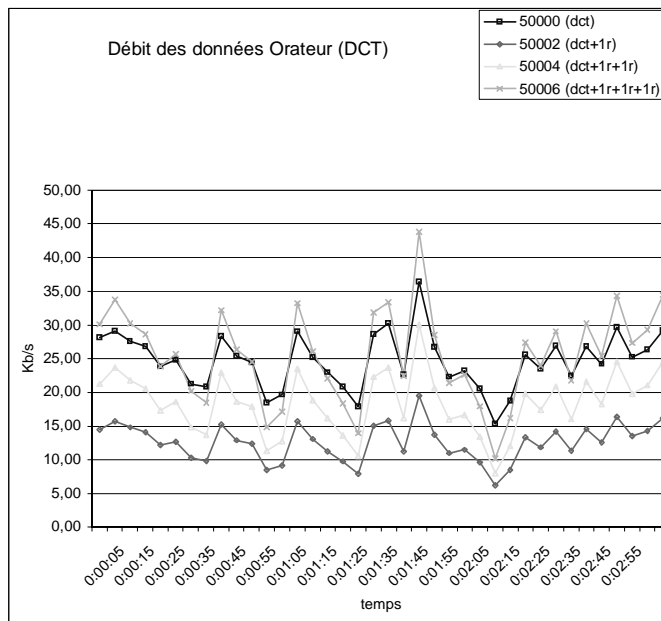


Figure 7. (a) débits de la luminance d'une vidéo moyennement changeante avec plusieurs flux de raffinement.

Les figures 8a et 8b montrent que si l'on additionne le débit de 3 passes de raffinement dans 3 sous-flux, on obtient le même résultat qu'en mettant les 3 raffinements dans le même sous-flux. Ainsi, on pourra découper les sous-flux en fonction des capacités des récepteurs. Ces remarques sont également vraies pour les flux de chrominance et ceux correspondants au codage en sous-bande.

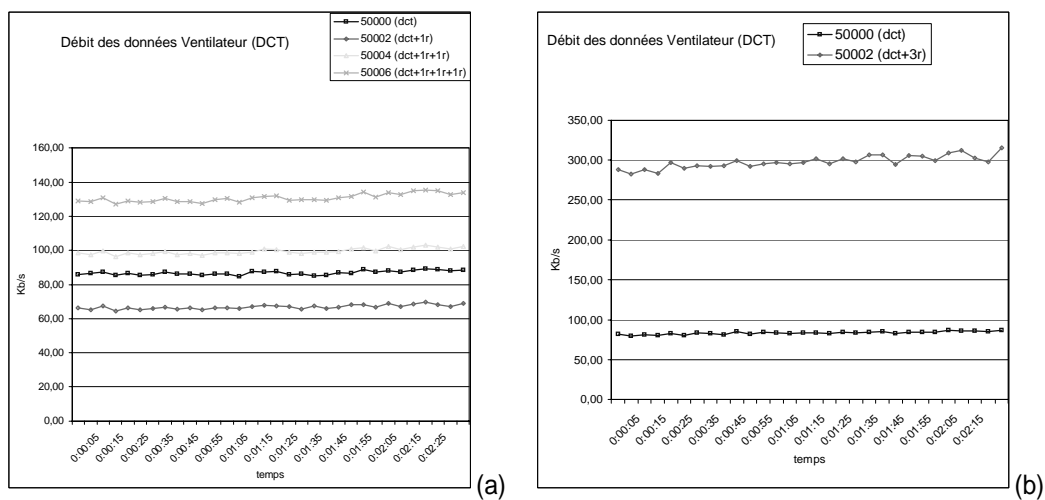


Figure 8.

- (a) débits de la luminance d'une vidéo très changeante avec plusieurs flux de raffinement
- (b) débits de la luminance d'une vidéo très changeante avec un seul flux de raffinement.

Nous avons également testé le codeur PVH sur un réseau IPv6[11]. Sur la figure 9, nous voyons que globalement, H261 génère moins de paquets, mais plus gros que PVH. Bien entendu, la taille des données par paquet ne change pas en fonction du protocole utilisé dans la couche réseau.

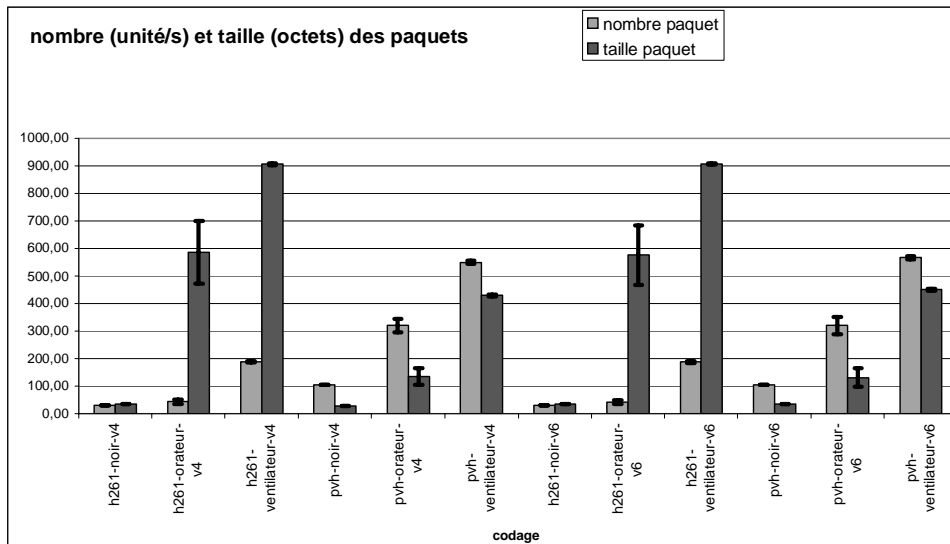


Figure 9. Comparaison de la taille et du nombre de paquets générés par H261 et PVH en IPv4 et en IPv6.

En revanche, ce qui est modifié, c'est la part de charge utile dans chacun de ces paquets. La figure 10, nous montre ces différences. Les barres noires verticales symbolisent l'écart-type des valeurs. La surcharge totale est obtenue par l'équation 3.

$$\text{Surcharge totale} = \frac{\text{nombre de bits d'entête en 5s}}{100}$$

Equation 3. Surcharge totale.

La surcharge par paquet est importante lorsqu'il y a peu de données à transmettre. Elle n'est plus très importante lorsque le débit correspond à une vidéoconférence. Par contre, étant donné que PVH a tendance à créer beaucoup de paquets de petite taille, les entêtes prennent une part non négligeable dans l'ensemble du trafic.

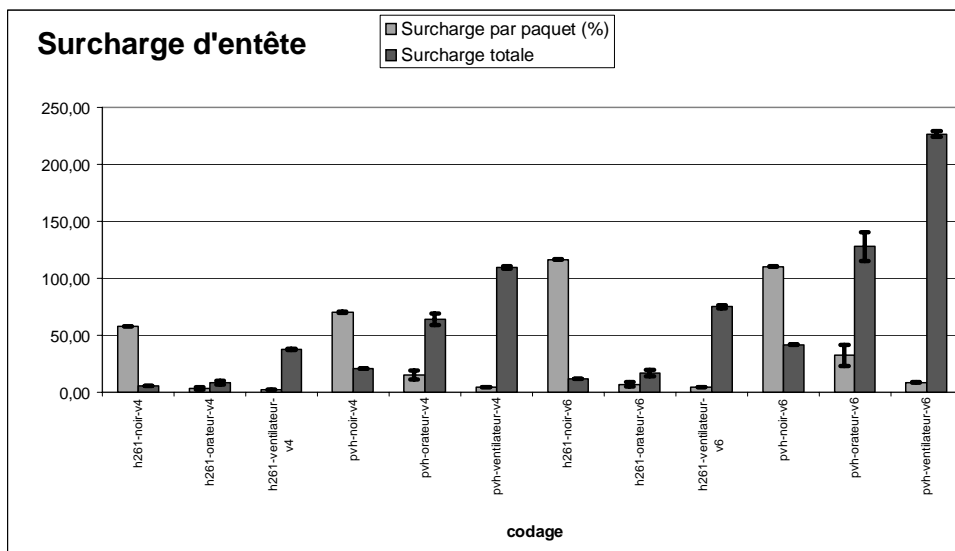


Figure 10. Surcharge due à l'entête des paquets H261 et PVH en IPv4 et en IPv6.

Toutes ces valeurs montrent que PVH est plutôt bien adapté à la diffusion de vidéo sur Internet. En effet, le débit minimum requis, c'est à dire le flux de base de la luminance, est inférieur à la capacité de la plupart des



modems actuels. De plus, il existe une règle qui veut que le débit des sessions multipoint soit inférieur à 128Kb/s. Or, PVH génère une vidéo de bonne qualité pour un débit inférieur à ce seuil. On peut même descendre le taux de rafraîchissement à environ 12 images/s ce qui reste très correct à regarder. Dans ce cas, le débit baisse d'à peu près 30 %.

## ■ Conclusion et perspectives

Cet article présente une évaluation de la diffusion de vidéo hiérarchique sur Internet avec un codage PVH. En premier lieu, nous avons effectué un rappel sur le codage et les algorithmes de traitement du signal.

Après avoir présenté PVH, nous avons évalué les flux générés par ce dernier. Nous les avons comparés à intra-H261 en IPv4 et IPv6. Cette expérimentation nous permet de dire que PVH est bien adapté à la diffusion de vidéoconférence sur Internet.

Il reste à étudier la manière de faire évoluer dynamiquement le volume associé aux flux pour s'adapter automatiquement aux capacités des récepteurs.

## ■ Bibliographie

1. Amir (A.), MacCanne (S.), Vetterli (M.). - A layered DCT coder for internet video. *IEEE international conference on image processing*, Lausanne, September 1996.

2. MacCanne (S.), Vetterli (M.), Jacobson (V.). - Low-complexity video coding for receiver-driven layered multicast. *IEEE journal on selected areas in communications*, 16, no 6, 983-1001, August 1997.

- Video Codec for Audiovisual Services at p \* 64kbits/s. *ITU-T Recommendation H.261*, 1993.

Waitzman(D.), Partridge(C.). - Distance Vector Multicast Routing Protocol. *RFC1075*, November 1988.

Deering (S.), Estrin (D.), Farinacci (D.), Jacobson (V.), Helmy (A.), Meyer (D.), Wei (L.). - Protocol Independent Multicast Version 2 Dense Mode Specification. *Internet Draft draft-ietf-pim-v2-dm-03.txt*, June 1999.

Moy (J.). - Multicast Extensions to OSPF. *RFC1584*, March 1994.

Ballardie (A.). - Core Based Trees (CBT) Multicast Routing Architecture. *RFC2201*, September 1997.

Estrin (D.), Farinacci (D.), Helmy (A.), Thaler (D.), Deering (S.), Handley (M.), Jacobson (V.), Liu (C.), Sharma (P.), Wei (L.). - Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification. *RFC2362*, June 1998.

Schulzrinne (H.), Casner (S.), Frederick(R.), Jacobson (V.). - RTP: a Transport Protocol for Real Time Applications. *RFC1889*, January 1996.

Shapiro (J.). - Embedded image coding using zerotrees of wavelet coefficients. *IEEE transactions on signal processing*, 41, no 12, 3445-3462, December 1993.

Deering (S.), Hinden (R.). - Internet Protocol, Version 6 (IPv6) Specification. *RFC2460*, December 1998.

*Mots clés : diffusion, vidéo, codage hiérarchique.*



