

Streaming vidéo adaptatif avec un modèle de distorsion récursif

Nicolas TIZON, Béatrice PESQUET-POPESCU

Telecom ParisTech, CNRS LTCI, Laboratoire Traitement du Signal et Images
46 rue Barrault, 75634 Paris Cedex 13, France

tizon@telecom-paristech.fr, pesquet@telecom-paristech.fr

Résumé – A partir d’une connaissance partielle de l’état du réseau, le problème posé est celui du choix d’une politique de transmission optimale afin de maximiser la qualité perçue par l’utilisateur pour des conditions de réception données. Le premier apport de cette étude réside dans la définition de modes de fonctionnement du serveur en fonction d’une estimation de la congestion. La stratégie d’adaptation s’appuie sur le codage scalable (Scalable Video Coding - SVC) et en particulier la scalabilité en qualité. Un modèle de distorsion récursif est proposé afin de hiérarchiser les paquets en attente en fonction de leur contribution à la distorsion totale. A partir de l’évaluation des capacités du réseau, les paquets sont alors transmis de façon à minimiser la distorsion mesurée au niveau du décodeur. Comparés aux systèmes de streaming classiques, les résultats expérimentaux montrent que notre approche permet de mieux faire face aux dégradations des conditions du réseau tels que la réduction de la bande passante ou l’augmentation du taux d’erreur.

Abstract – This paper proposes a video streaming system optimizing resource utilization when the media server only disposes of long term feedbacks from the client. We developed a scheduling algorithm based on the scalable video coding (SVC) and more particularly on the SNR scalability. We also propose a recursive distortion model that allows to differentiate queueing packets according to their contribution on the global distortion. From a network capacity evaluation, packets are transmitted in a way that minimizes the distortion measured at decoder side. Compared to more conventional streaming systems, experimental results show that our approach allows to better face network condition degradation like bandwidth reduction or packet error rate increase.

1 Introduction

L’adaptation de débit est un problème crucial pour les applications de streaming mettant en oeuvre des réseaux à bande passante limitée et susceptibles de perdre des paquets. Dans une grande majorité des cas pratiques, le serveur de médias se trouve très en amont du réseau et des couches protocolaires propres à la radio, dans le cas d’un réseau sans fil, rendant ainsi très difficile l’estimation en temps réel de l’état du canal. A notre connaissance, [1] constitue l’étude la plus complète visant à contrôler l’envoi des données en fonction des rapports RTCP. Les différents champs d’un RTCP-RR (Receiver Report) [2] sont utilisés afin de connaître le niveau de remplissage de la file d’attente en entrée du canal sans fil. Dans cette étude, nous utiliserons de façon similaire les RTCP-RR afin d’évaluer la congestion dans le réseau et la capacité du canal.

L’optimisation suivant un critère de débit-distorsion (RaDiO) de la politique de transmission des paquets d’un flux vidéo a fait l’objet d’un article [3] à la base de nombreux travaux plus récents comme [4] où la contrainte n’est plus le débit mais la congestion. En exploitant les apports du codage vidéo scalable et la définition de régions d’intérêts (ROI), nous avons proposé dans [5] un système de streaming adaptatif s’appuyant sur le protocole radio et une définition simple des régions. La solution apportée vise à maximiser la qualité perçue par le client en luttant contre la congestion causée par la capacité d’un paquet à surcharger le réseau. Ce point de vue est assez proche de celui

proposé en [4]. Dans [5], la décision de transmettre un paquet se fait en tenant compte de l’impact sur l’envoi de paquets plus prioritaires en attente.

Les solutions de type “RaDiO” s’appuient sur un calcul relativement complexe de la distorsion engendrée par les pertes de paquets. En effet cette méthode nécessite de décoder les différentes combinaisons de paquets correspondant aux différentes politiques d’envoi envisageables. Dans le cadre du codage vidéo distribué [6], propose un modèle de distorsion récursif afin d’optimiser le choix d’un schéma de prédiction au décodeur. Dans cette étude, nous nous appuyons sur ce modèle pour évaluer l’augmentation de distorsion provoquée par la perte d’un paquet dans le cadre du codage scalable SVC.

Dans la partie suivante nous présentons les aspects protocolaires permettant de contrôler le délais de transmission et le niveau de congestion. Nous présentons ensuite le modèle de distorsion récursif ainsi que l’algorithme de transmission des paquets qui en découle. Après avoir présenté les résultats expérimentaux permettant de valider l’approche, nous concluons.

2 Détection de la congestion et contrôle de débit

L’idée principale de notre méthode est de maintenir la continuité de l’affichage au niveau du récepteur en évitant la congestion dans le réseau. En les recoupant avec l’historique des pa-

vante :

$$D_X^q = \alpha_X^q (\sigma_X^q)^2 2^{-2R_X^q}. \quad (3)$$

Dans [6], cette formule permet de calculer récursivement la distorsion des images biprédites (type B). Ainsi pour B2 :

$$\begin{aligned} D_{B2}^q &= \alpha_{B2}^q (\sigma_{B2}^q)^2 2^{-2R_{B2}^q} \\ &= \alpha_{B2}^q \left(M_{2,2}^q + \frac{1}{4} D_I + \frac{1}{4} D_P \right) 2^{-2R_{B2}^q}, \end{aligned} \quad (4)$$

où $M_{2,2}^q$ représente l'erreur de prédiction lorsque l'estimation de mouvement est faite par rapport à l'image originale et R_{B2}^q est le débit des données correspondantes.

Par la suite, nous supposons que chaque paquet de la couche de base ($q = 0$) est bien reçu à temps par le décodeur et qu'un paquet n'est envoyé que si les paquets dont il dépend pour être décodé ont déjà été envoyés. L'algorithme proposé doit donc arbitrer l'envoi des différents paquets de la couche de raffinement ($q = 1$). Notons ΔD_X la différence entre la distorsion de la vidéo lorsque le paquet de la couche de raffinement de l'image X est perdu (ou non transmis) et la distorsion de la vidéo lorsqu'il est reçu. Cette distorsion, calculée par le serveur avant l'envoi du paquet, dépend des paquets déjà envoyés et censés être reçus par le client. Pour l'image I et avec un taille de GOP de 2, on a :

$$\Delta D_I = D_I^1 - D_I^0 + \frac{1}{4} \alpha_{B2}^q (D_I^1 - D_I^0) 2^{-2R_{B2}^q}, \quad (5)$$

où q prend la valeur 1 si le paquet de la couche de raffinement de l'image B2 a déjà été envoyé et la valeur 0 sinon. Dans Eq. 5, les éléments d'indice P correspondants à la première image du second GOP n'apparaissent pas. En effet, le schéma de codage mis en œuvre utilise le concept d'images clé (fond grisé sur la figure). Ces images sont encodées en utilisant une boucle de prédiction s'appuyant uniquement sur les images de qualité basse. La distorsion de ces images n'est conditionnée que par la réception ou la perte de leur couche de raffinement. L'équation Eq. 5 peut être généralisée pour les autres paquets avec d'autres tailles de GOP. Par ailleurs, avec une scalabilité SNR de type FGS, l'information de mouvement utilisée pour la prédiction est la même pour la couche de base et pour les couches de raffinement. On a donc $M_{i,i}^0 = M_{i,i}^1$ pour toutes les images d'un GOP. De même, on suppose que le coefficient de normalisation α est constant pour une image, quelque soit la couche de scalabilité SNR considérée ($\alpha_X^0 = \alpha_X^1$). Ainsi la relation entre les distorsions d'une image décodée avec et sans la couche de raffinement est donnée par :

$$D_X^1 = D_X^0 2^{-2R_X^q}. \quad (6)$$

De façon générale, la distorsion d'une image bi-prédite peut s'exprimer en fonction des distorsions D_{ref0} et D_{ref1} des deux images de référence :

$$D_B^0 = a_B^0 (D_{ref0} + D_{ref1}) + b_B^0, \quad (7)$$

où a_B^0 et b_B^0 sont deux constantes fixées de façon empirique. De plus, à partir de la relation $D_B^1 = D_B^0 2^{-2R_B^1}$ on peut définir les constantes a_B^1 et b_B^1 telles que : $a_B^1 = a_B^0 2^{-2R_B^1}$ et

$b_B^1 = b_B^0 2^{-2R_B^1}$. A partir de ces relations, on peut exprimer la diminution de distorsion associée aux deux types d'images : image clé (indice K) et bi-prédite (indice B), lorsque le paquet de la couche de raffinement est reçu au décodeur :

$$\begin{aligned} \Delta D_K &= (D_K^0 - D_K^1) \left(1 + \sum a_{B_i}^{q_i} \right), \\ \Delta D_B &= D_B^0 \left(1 - 2^{-2R_B^1} \right) \left(1 + \sum a_{B_i}^{q_i} \right), \end{aligned} \quad (8)$$

où les $a_{B_i}^{q_i}$ sont les constantes, introduites dans Eq. 7, des images bi-prédites pour lesquelles l'image considérée sert de référence. Les variables q_i prennent la valeur 1 lorsque la couche de raffinement de l'image a été envoyée et la valeur 0 sinon.

Les paquets de la couche de raffinement en attente dans le serveur sont examinés périodiquement en vue d'être transmis, mis en attente ou supprimés de la file d'attente. Cette prise de décision se fait de façon périodique à la fréquence de rafraîchissement maximale de la vidéo. Avant d'envisager l'envoi éventuel d'un paquet, l'algorithme s'assure que sa date limite de réception n'est pas passée et qu'une proportion suffisante du délai point à point maximum est encore disponible avant d'atteindre cette date limite. Si ce n'est pas le cas, le paquet est supprimé de la file d'attente et on examine alors le paquet suivant de la file dans l'ordre croissant des instants d'échantillonnage. Notons n l'indice de cette image. On note $\Delta D(n)$ la distorsion associée à ce paquet. On construit alors un paquet "virtuel" en concaténant les paquets en attente, appartenant aux images d'indices n_0, \dots, n_N et tels que : $\forall k = n_0, \dots, n_N, \Delta D(k) > \Delta D(n)$. La date limite de réception de ce paquet est donnée par le paquet ayant l'instant d'échantillonnage le plus ancien de l'ensemble. A partir des paramètres du réseau définis partie 2, le serveur peut estimer à chaque envoi d'un paquet le niveau de remplissage du buffer en entrée du lien sans fil. A partir de ce niveau de remplissage et de l'estimation de la capacité du canal R_t , le serveur s'assure que l'envoi du paquet d'indice n n'induit pas une réception du paquet virtuel après sa date limite. Si c'est le cas, le paquet d'indice n est envoyé. Sinon, le paquet est mis en attente et on examine le paquet suivant de la file dans l'ordre croissant des instants d'échantillonnage. Les distorsions ΔD associées à chaque paquet sont mises à jour en fonction des paquets qui ont été envoyés.

Dans la partie suivante, nous présentons des résultats de simulation qui permettent de valider le modèle de distorsion ainsi que l'algorithme de transmission qui vient d'être décrit.

4 Résultats expérimentaux

Pour évaluer notre approche, nous simulons un scénario de transmission dans lequel les conditions du réseau se dégradent progressivement au cours du temps. La vidéo utilisée d'une durée de 75 secondes est obtenue en concaténant les séquences : costguard, hall, mobile, stefan, bus, city, container et foreman à la résolution QCIF et à la fréquence 15Hz. Cette vidéo est encodée avec une couche de raffinement SNR (FGS) et une taille de GOP de 2 dans un premier temps, afin de valider le modèle de

distorsion. Le facteur de quantification est fixe et donc le débit varie fortement dans le temps suivant le contenu. On observe Fig. 3 que la distorsion des images de haute qualité, calculée en fonction de la distorsion des images de basse qualité et des débits des couches de raffinement est très proche de la distorsion mesurée. Par ailleurs, on observe une bonne adéquation du modèle en fixant $a_B^0 = 0.5$ et $b_B^0 = 0$ pour l'ensemble des images bi-prédites de la séquence.

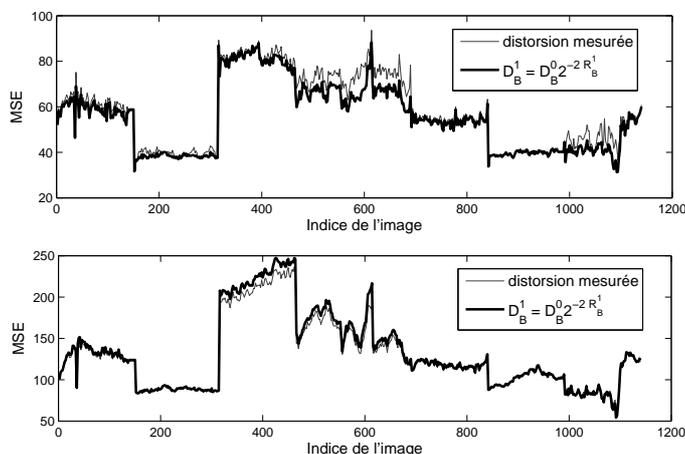


FIG. 3 – Estimation de la distorsion D_B^1 des images bi-prédites en fonction de la distorsion D_B^0 et du débit R_B^1 pour un GOP de taille 2 (haut : $D_K = D_K^1$, bas : $D_K = D_K^0$)

Par la suite, la taille de GOP utilisée est de 8. Au début de la session, le canal est établi avec une bande passante de 128kbps et un taux d'erreur par blocs de 3,3%. Autour de la 30^e seconde (450^e image), le taux d'erreur atteint 10% et la bande passante chute à 100kbps. Dans un deuxième temps, le taux d'erreur reste stable à 10% et la bande passante atteint un minimum de 80kbps. Les rapports RTCP-RR sont reçus par le serveur avec une période de une seconde et le délai initial d'affichage des images est fixé à 5 secondes. Un compromis entre stabilité et rapidité d'adaptation est trouvé en fixant les paramètres $\alpha_{max} = 100\%$ et $T_{congest} = 1.5s$.

On compare alors l'évolution du PSNR de la vidéo décodée pour trois politiques de transmission différentes : aucun contrôle, simple contrôle sur les dates limites de réception et le contrôle de congestion optimisé précédemment décrit. On ob-

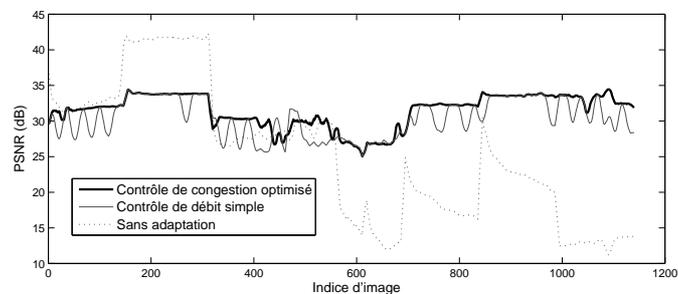


FIG. 4 – Evolution du PSNR dans le temps (QCIF@15Hz).

serve Fig. 4 la capacité de notre algorithme à faire face à de très fortes détériorations de l'état du réseau. De plus la prise en compte des paquets de plus haute importance lors de l'envoi d'un paquet permet d'utiliser plus efficacement la ressource, en particulier pendant les états de transition. Le gain de PSNR moyen sur la séquence entière est de 1.13dB par rapport à une méthode de contrôle de débit simple qui ne hiérarchise pas les paquets en fonction de leur importance.

méthode de transmission	PSNR moyen
Contrôle de congestion optimisé	31.48 dB
Contrôle de débit simple	30.35 dB
Sans adaptation	25.32 dB

TAB. 1 – PSNR moyen pour différentes méthodes de transmission.

5 Conclusion

Ce papier propose un schéma complet de streaming vidéo optimisé pour les réseaux sans fil. La solution apportée se focalise sur le cas où le serveur reçoit uniquement des rapports RTCP-RR comme information de retour du réseau. Les conditions de réception du client sont estimées en distinguant deux modes de fonctionnement : congestionné et non-congestionné. Nous présentons également un modèle de distorsion récursif qui permet de calculer de façon simple la diminution de distorsion attendue lors de l'envoi d'un paquet contenant des informations de raffinement en qualité. La hiérarchisation des paquets ainsi obtenue permet de mettre en œuvre une stratégie de transmission qui minimise le risque qu'un paquet envoyé pénalise la transmission de paquets plus importants. Les travaux futurs s'intéresseront à une généralisation de la solution dans le cas d'un canal partagé entre plusieurs utilisateurs.

Références

- [1] N. Baldo, U. Horn, M. Kampmann, and F. Hartung, "RTCP feedback based transmission rate control for 3G wireless multimedia streaming," *IEEE Int. Symp. Personal, Indoor and Mobile Radio Com.*, vol. 3, pp. 1817–1821 Vol.3, Sept. 2004.
- [2] *Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback*, RFC4585.
- [3] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *Microsoft Corporation, Tech. Rep. MSR-TR-2001-35*, Feb. 2001.
- [4] E. Setton and B. Girod, "Congestion-distortion optimized scheduling of video over a bottleneck link," *IEEE Workshop on Multimedia Signal Proc.*, pp. 179–182, Siena, Italy, Sept. 2004.
- [5] N. Tizon and B. Pesquet-Popescu, "Scalable and media aware adaptive video streaming over wireless networks," *EURASIP Journal on Advances in Signal Processing*, 2008.
- [6] T. Maugey, T. André, B. Pesquet-Popescu, and J. Farah, "Analysis of error propagation due to frame losses in a distributed video coding system," *EUSIPCO*, Lausanne, Aug. 2008.