

Évaluation subjective des effets de pertes de paquets sur des vidéos codées en H.264/AVC

Fadi Boulos¹, David Hands², Benoît Parrein¹ et Patrick Le Callet¹

¹Université de Nantes, Nantes Atlantique Universités
IRCCyN CNRS UMR 6597, Polytech'Nantes, rue Christian Pauc, 44306 Nantes, France

²BT Innovate, Martlesham Heath, Ipswich, UK

email: fadi.boulos@univ-nantes.fr

Résumé

La diffusion de vidéos sur Internet connaît un grand essor depuis quelques années. Pourtant, la qualité d'usage associée à ce service varie énormément. Ceci est essentiellement dû aux restrictions de bande passante et à la fluctuation de la qualité de service. Pour mieux comprendre comment l'utilisateur final réagit face à une qualité d'usage variable, nous proposons dans ce papier d'analyser les effets perceptuels des pertes de paquets. Pour ce faire, nous présentons les résultats obtenus lors d'un test subjectif d'évaluation de la qualité de vidéos codées en H.264/AVC et ayant subi différents taux et motifs de pertes de paquets. Nous en tirons des conclusions intéressantes concernant l'impact que peuvent avoir les paramètres liés au réseau et ceux liés au contenu de la séquence sur la qualité perceptuelle. En particulier, nous quantifions l'importance de la distribution des pertes et de la position du changement de scène dans la séquence vidéo.

Mots clefs

Qualité d'usage, codage H.264/AVC, perte de paquets, tests subjectifs.

1 Introduction

La qualité d'usage est un facteur important du point de vue d'un Fournisseur d'Accès à Internet (FAI) car c'est une indication du niveau de satisfaction des clients. Mais mesurer cette qualité d'usage est une tâche complexe. Actuellement, la méthode la plus fiable d'évaluation de la qualité d'usage d'un service vidéo est la conduite de tests subjectifs de qualité. Ces tests consistent en un groupe d'observateurs qui ont comme la tâche l'évaluation de la qualité de séquences vidéos généralement dégradées par le codage, les erreurs de transmission ou les post-traitements.

Trois paramètres essentiels définissent la qualité de service d'un réseau : le taux de pertes de paquets, le délai et la variation de délai (gigue). Dans le cadre de ce travail, nous supposons que ces trois paramètres peuvent être réduits à la perte de paquets, au vu de leurs conséquences indirectes.

Cette hypothèse est vraie dans le cas où le délai d'attente est court, forçant par exemple un routeur à éliminer tout paquet arrivant après la fin du délai.

Dans ce papier, nous passons tout d'abord en revue quelques travaux de la littérature traitant de la problématique de la qualité perceptuelle de vidéos dégradées par des pertes de paquets. Ensuite, nous expliquons la méthodologie de travail adoptée ; nous décrivons ainsi les processus de simulation de pertes et de mise en place des tests subjectifs d'évaluation de qualité. Enfin, nous présentons et discutons les résultats obtenus tout en soulignant les perspectives pour les travaux futurs.

2 Etat de l'art

L'évaluation de la qualité de vidéos dégradées par les pertes de paquets reste une problématique ouverte. Nous présentons brièvement dans cette section quelques travaux de la littérature étudiant l'un ou l'autre des aspects de cette problématique.

Kanumuri *et al.* examinent le seuil de visibilité des dégradations liées aux pertes de paquets dans [1, 2]. En vue du développement d'un modèle objectif linéaire pour le suivi de la qualité de vidéos sur un réseau IP, des tests subjectifs sont conduits pour quantifier l'impact des pertes sur la qualité visuelle. Durant ces tests, les observateurs doivent réagir quand ils perçoivent une dégradation de qualité dans des séquences vidéos codées en MPEG-2 [1]. Il est ainsi montré que des facteurs comme la quantité de mouvement présente dans les paquets perdus, le nombre de paquets perdus et le type des images contenues dans ces paquets influencent beaucoup la visibilité des dégradations. Des résultats similaires sont obtenus dans [2] pour des vidéos codées en H.264/AVC mais avec une différence : la quantité de mouvement n'est plus à considérer dès lors qu'une compensation de pertes basée mouvement est appliquée.

Dans [3], Reibman et Poole étudient les effets des caractéristiques du contenu des séquences sur la visibilité des dégradations. Les auteurs concluent que cette visibilité augmente si la perte a lieu au changement de scène et diminue

si elle est située juste avant ou après. Ils démontrent aussi que le mouvement de la caméra (zoom par exemple) peut augmenter la visibilité des pertes.

D'autre part, plusieurs études ont examiné la relation entre la longueur de la rafale de paquets perdus et la qualité de la vidéo. Dans [4], Liang *et al.* montrent que la dégradation résultante d'une perte de paquets dépend fortement du modèle de pertes ; ainsi, la perte d'un nombre de paquets consécutifs produit une dégradation plus importante que celle produite par le même nombre de paquets perdus isolément. À noter que la mesure de distorsion utilisée est l'erreur quadratique moyenne et non les tests subjectifs. Dans [5], des tests subjectifs comprenant des vidéos au format Standard Definition (SD) pour des applications de TV sur IP montrent que pour des pertes de moins de 200 ms, la qualité d'usage est plus influencée par le nombre de pertes que par leur longueur. Ceci implique que pour un même nombre de paquets perdus, la qualité est meilleure dans le cas d'une perte de plusieurs paquets consécutifs que dans le cas de plusieurs pertes singulières. La contradiction de ces résultats avec ceux de [4] nous a motivés à étudier de plus près l'effet de la distribution des pertes sur la qualité visuelle.

Le travail présenté dans ce papier vise à mieux comprendre la réaction des observateurs vis-à-vis des dégradations spatio-temporelles que peut subir une séquence vidéo. Plus particulièrement, nous essayons de répondre aux deux questions suivantes :

1. Comment le motif de pertes contribue-t-il à la dégradation globale ?
2. A quel point le contenu de la séquence influence-t-il la note de qualité ?

3 Simulation de pertes de paquets et tests subjectifs

Dans la norme de codage H.264/AVC [6], chaque "slice" de données codées est encapsulée dans une unité NALU (Network Abstraction Layer Unit). Nous choisissons de simuler les pertes de paquets au niveau NAL dans le flux binaire car nous pouvons ainsi contrôler les positions spatiale et temporelle de la perte. Ces dernières correspondent respectivement à la position de la "slice" dans la séquence de NALUs du flux binaire et au numéro de l'image. À noter que la position spatiale est approximativement déterminée car une "slice" peut contenir une ou plusieurs rangées de macroblocks.

Nous fixons la taille maximale d'une "slice" à 1450 octets, en deçà de la taille maximale d'une unité de transport sur Ethernet (1500 octets). Les octets restants (au nombre de 50) servent à l'en-tête RTP/UDP/IP (40 octets) et aux éventuels octets additionnels utilisés par le codeur au-delà de la limite fixée. Ainsi, chaque NALU est encapsulée dans un paquet IP ce qui rend notre simulation réaliste car le taux de pertes au niveau NAL correspond au taux de pertes au

niveau application (RTP par exemple). Dans la suite du papier, les termes NALU et "paquet" seront confondus.

Le simulateur de pertes utilisé dans ce travail est basé sur [7]. Nous choisissons plusieurs motifs de pertes où longueur de rafale et nombre de pertes (et par suite taux de pertes) sont variés. Un résumé de tous les motifs de pertes utilisés dans le test subjectif est donné dans le tableau 1. Les pertes sont généralement introduites dans les images I et un test préliminaire vérifie que les séquences obtenues couvrent l'intégralité de l'échelle de qualité.

Dans le tableau 1, la fréquence des pertes indique leur nombre d'occurrences toutes tailles de rafale confondues. La taille d'une rafale représente le nombre de paquets consécutifs perdus. Le pourcentage de paquets perdus est équivalent au rapport (multiplié par 100) du nombre de paquets perdus de la séquence sur le nombre total de paquets dans la séquence. Par exemple, pour une séquence codée à 1,5 Mbit/s, 4 pertes de paquets de longueur 20 paquets consécutifs chacune engendrent une perte de 1,1% du nombre total de paquets dans la séquence.

Pour éviter que la position temporelle de la perte influence le jugement des observateurs, nous avons conservé les mêmes positions pour toutes les séquences. Ainsi, les pertes uniques ont toujours eu lieu à la sixième seconde de la séquence vidéo, indépendamment de la taille de la rafale. Pour les motifs de pertes à plusieurs occurrences (4 et 8), la séquence temporelle pendant laquelle les pertes sont survenues se situe entre la quatrième et la huitième seconde. L'intervalle de temps séparant les quatre et huit pertes a été fixé respectivement à une et une demi-seconde.

Le test subjectif a été conduit selon la recommandation BT.500-11 [8] de l'ITU (International Telecommunication Union). Pendant le test, les observateurs devaient noter la qualité de chaque séquence vidéo affichée sur un écran CRT à la fréquence de 25 images par seconde. Huit séquences vidéos (de dix secondes chacune) ont été utilisées pour le test. Ces séquences, propriété de BT, comprenaient des contenus à quantités de mouvement et niveaux de texture différents. La résolution utilisée était la SD (720 × 576) ; la structure du GOP était de la forme IBBPBBP... et sa longueur 24 images. Les codeur et décodeur de BT ont été utilisés. Un algorithme de compensation d'erreurs non-standard, implémenté dans le décodeur, était appelé à chaque fois qu'une partie du flux était perdue (à cause des pertes de paquets). Cet algorithme cherchait le meilleur macroblock spatio-temporel voisin pour remplacer le macroblock perdu. Aucune fonction de résistance aux pertes de paquets offerte par le standard H.264/AVC (par exemple Flexible Macroblock Ordering et Data Partitioning) n'a été utilisée dans ce test. Ce choix est justifié par la volonté d'évaluer la qualité de séquences vidéos ayant subi un codage classique. Par ailleurs, une étude des performances du couplage de FMO avec les régions d'intérêt est en cours.

TAB. 1 – Les motifs de pertes utilisés.

Débit (Mbit/s)	1,5								4									
Fréquence des pertes (par séquence)	1				4				8				1			4		
Taille de la rafale (paquets)	4	10	20	40	4	10	20	20	4	10	20	20	4	10	20	4	10	20
Pourcentage de paquets perdus	0,2	0,5	1,1	2,2	0,9	2,2	4,5	8,9	0,1	0,2	0,4	0,4	0,4	0,9	1,8			

La méthode d'évaluation par catégories absolues (à simple stimulus) a été utilisée pour recueillir les jugements des observateurs, les MOS (Mean Opinion Score). La figure 1 représente l'échelle de cinq catégories utilisée. 16 observateurs non experts ont participé au test dont la durée moyenne était de 35 minutes. Les 14 motifs de pertes du tableau 1 ont été appliqués aux séquences vidéos résultant ainsi en un total de 112 séquences. Nous avons intégré les séquences codées aux deux débits et n'ayant pas subi de pertes de paquets à l'ensemble de vidéos de test portant ainsi le total à 128 séquences.

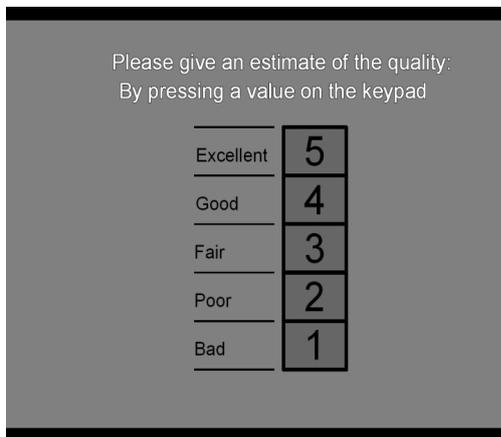


FIG. 1 – L'échelle à cinq catégories utilisée dans le test subjectif.

4 Résultats et discussion

Les résultats obtenus peuvent être classés dans deux catégories : les résultats liés au contenu de la séquence et ceux liés à la transmission sur le réseau.

Dans la première catégorie, nous montrons que pour N ($N \leq 40$) paquets consécutifs perdus dans des images I et B juste avant un changement de scène, la qualité visuelle reste égale à la qualité de la même séquence n'ayant subi aucune perte. Par contre, la perte d'une partie de la première image d'une nouvelle scène peut engendrer une diminution catastrophique de la qualité. Les courbes de la figure 2, représentant les valeurs des MOS pour les versions des séquences avec et sans pertes illustrent ces résultats. Les séquences ont été codées à un débit de 1,5 Mbit/s.

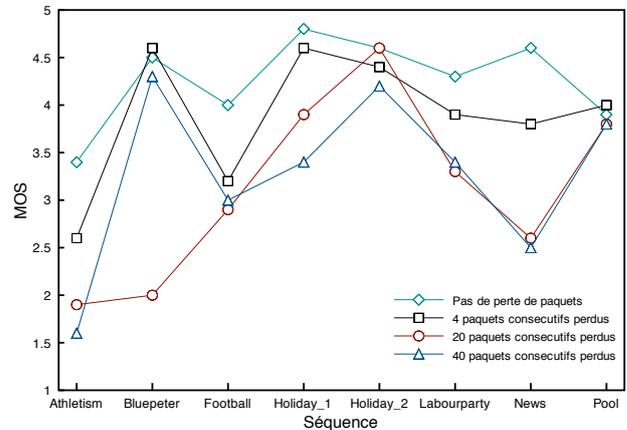


FIG. 2 – L'impact du changement de scène sur la qualité perceptuelle.

Les séquences contenant un changement de scène sont *Bluepeter*, *Holiday_2* et *Pool*. Pour les séquences sans changement de scène, nous remarquons une diminution de la qualité à mesure que le nombre de paquets perdus augmente. Par contre, ces trois séquences n'ont pas cette attitude.

Généralement, la première image d'une nouvelle scène est codée en image I même si elle n'est pas la première image du GOP. En observant les courbes des séquences *Holiday_2* et *Pool*, nous pouvons constater que la variation de MOS entre les différents scénarios de pertes est minimale sachant que le taux de pertes augmente de 0,22% à 2,2%. Ceci est dû au fait que la propagation temporelle de l'erreur causée par le processus d'inter-prédiction est stoppée quand une image I est rencontrée. Dans les deux cas susmentionnés, les pertes ont lieu dans l'image I située juste avant le changement de scène ce qui empêche l'erreur de se propager vers un nouveau GOP.

Dans la séquence *Bluepeter*, nous remarquons la même attitude pour tous les scénarios de pertes sauf celui où les pertes ont une longueur de 20 paquets. En effet, dans ce dernier cas, la perte a lieu dans la première image d'une nouvelle scène (une image I) ce qui fait que le décodeur est incapable d'appliquer une compensation de pertes efficace à cause de la perte des seules informations utiles. La figure 2 montre aussi que les versions de *Bluepeter*, *Holiday_2* et *Pool* avec pertes juste avant le changement de scène ont une diminution de moins de 0,5 MOS comparées à la version sans pertes. Ceci confirme la robustesse des séquences contenant un changement de scène face aux pertes de paquets.

D'autre part, nous étudions l'impact que peut avoir la distribution des pertes de paquets sur la note de qualité attribuée par un observateur à une séquence vidéo. Nous avons ainsi comparé deux scénarios de pertes différents mais qui accusent un même taux de pertes global (2,2%). Nous avons choisi les motifs suivants : perte unique de 40 paquets consécutifs et quatre pertes régulières de 10 paquets chacune. Les résultats sont présentés dans la figure 3.

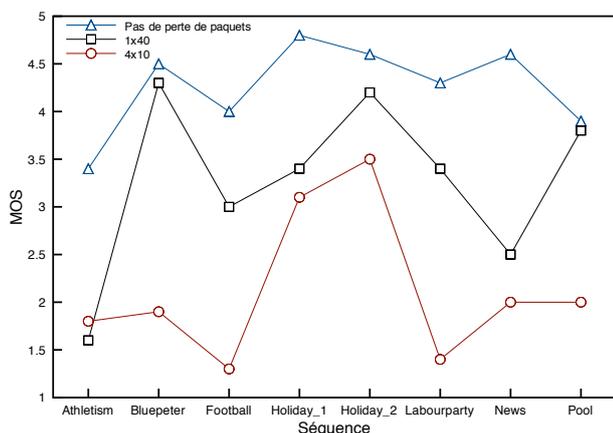


FIG. 3 – L'impact de la distribution des pertes sur la qualité perceptuelle.

Nous pouvons déduire des courbes de la figure 3 que la qualité d'une vidéo est généralement meilleure dans le cas d'une seule perte importante que dans le cas de plusieurs pertes moins importantes. Ceci est dû à la durée de propagation de l'erreur. En effet, dans la première situation, les pertes ont lieu dans l'image I et les images B et P suivantes mais la propagation reste confinée à l'intérieur du GOP concerné tandis que dans la seconde situation, l'erreur se propage à plusieurs images à chaque perte ; et comme les pertes ont lieu dans quatre images I consécutives, la propagation temporelle peut durer jusqu'à quatre secondes (40% de la durée totale de la séquence). À noter que la qualité de la séquence *Athletism* qui contient une grande quantité de mouvement est mauvaise pour les deux motifs de pertes car l'algorithme de compensation de pertes n'a pas de bonnes performances à ce taux de pertes.

5 Conclusion et perspectives

Nous avons présenté dans ce papier notre travail visant une compréhension plus profonde de la qualité d'usage des services vidéos. Nous avons ainsi décrit la méthodologie d'évaluation subjective de qualité et l'introduction de pertes de paquets dans le flux binaire des séquences en utilisant pour cette dernière la couche réseau du codage H.264/AVC. Nous avons aussi établi un lien entre le modèle des pertes subies pendant la transmission et la qualité perceptuelle résultante de la vidéo. De plus, nous avons

identifié l'importance de l'effet de masquage de l'erreur que peut causer un changement de scène dans une séquence vidéo.

Le présent travail peut être exploité dans deux contextes : le premier contexte est celui d'un critère objectif de qualité spécifique aux dégradations subies lors de la transmission tandis que le second s'insère dans la problématique du codage conjoint source-canal. Des recherches sont en cours pour valider l'efficacité de notre approche dans ces deux contextes.

Références

- [1] S. Kanumuri, P.C. Cosman, et A.R. Reibman. "A Generalized Linear Model for MPEG-2 Packet-Loss Visibility". Dans *14th International Packet Video Workshop*, Irvine, Décembre 2004.
- [2] S. Kanumuri, S.G. Subramanian, P.C. Cosman, et A.R. Reibman. "Predicting H.264 Packet Loss Visibility using a Generalized Linear Model". Dans *Proceedings of IEEE International Conference on Image Processing, ICIP*, pages 2245–2248, Octobre 2006.
- [3] A.R. Reibman et D. Poole. "Predicting Packet-Loss Visibility Using Scene Characteristics". Dans *16th International Packet Video Workshop*, Lausanne, Novembre 2007.
- [4] Y. Liang, J. G. Apostolopoulos, et B. Girod. "Analysis of Packet Loss for Compressed Video : Does Burst-Length Matter ?". Dans *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Avril 2003.
- [5] Architecture & Transport Working Group. Triple-play Services Quality of Experience (QoE) Requirements. Rapport technique TR-126, DSL Forum, Décembre 2006.
- [6] Joint Video Team (JVT) of ISO/IEC MPEG et ITU-T VCEG. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC). JVT-G050r1, Mai 2003.
- [7] Joint Video Team (JVT) of ISO/IEC MPEG et ITU-T VCEG. SVC/AVC Loss Simulator. JVT-Q069, Octobre 2005. Disponible sur http://wftp3.itu.int/av-arch/jvt-site/2005_10_Nice/.
- [8] International Telecommunication Union-Radiocommunication Sector. Methodology for the Subjective Assessment of the Quality of Television Pictures. ITU-R BT.500-11, Juin 2002.