

RFC 8761 : Video Codec Requirements and Evaluation Methodology

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 17 avril 2020

Date de publication du RFC : Avril 2020

<https://www.bortzmeyer.org/8761.html>

Si vous passez toutes vos soirées avachi[Caractère Unicode non montré ¹] Je devant un écran à regarder des séries sur Netflix, ce RFC est pour vous. Il est le cahier des charges d'un futur codec vidéo libre pour l'Internet. Ce codec devra gérer beaucoup de cas d'usage différents.

Concevoir un codec vidéo n'est jamais facile. Il y a beaucoup d'exigences contradictoires, par exemple minimiser le débit en comprimant, tout en n'exigeant pas de la part du processeur trop de calculs. Ce RFC, le premier du groupe de travail netvc <<https://datatracker.ietf.org/wg/netvc/>> de l'IETF, se limite au cas d'un codec utilisé au-dessus de l'Internet, mais cela fait encore plein d'usages différents, du "streaming" à la vidéoconférence en passant par la vidéosurveillance (celle que les politiciens hypocrites appellent la vidéoprotection, mais le RFC est plus sincère et utilise le bon terme.)

Le problème est d'autant plus difficile que plein de choses peuvent aller mal sur l'Internet, les paquets peuvent être perdus, les bits modifiés en route, etc. En pratique, tous ces incidents n'ont pas la même probabilité : les pertes sont bien plus fréquentes que les corruptions, par exemple. Si on n'est pas trop pressés (cas du téléchargement), TCP ou d'autres protocoles de transport similaires fournissent une bonne solution. Si on a des exigences temporelles plus fortes (diffusion en direct, « temps réel »), il faut parfois renoncer à TCP et, soit utiliser des réseaux idéaux ("*managed networks*", dit le RFC, qui reprend le terme marketing de certains opérateurs qui tentent de nous faire croire qu'il existerait des réseaux sans pertes), soit trouver d'autres moyens de rattraper les pertes et les erreurs.

Depuis qu'on fait de la vidéo sur l'Internet, c'est-à-dire depuis très longtemps (malgré les prédictions des opérateurs de télécommunications traditionnels qui étaient unanimes, dans les années 1990, pour dire que cela ne serait jamais possible), toute une terminologie a été développée pour communiquer sur

1. Car trop difficile à faire afficher par L^AT_EX

ce sujet. La section 2 de notre RFC rappelle les sigles importants, et quelques termes à garder en tête. Ainsi, la « compression visuellement sans perte » désigne les méthodes de compression avec perte, mais dont on ne voit pas d'effet secondaire à l'œil nu.

La section 3 de notre RFC fait ensuite le tour des applications de la vidéo sur l'Internet, pour cerner de plus près leurs caractéristiques et leurs exigences propres. (Le cahier des charges proprement dit sera en section 4.) Chaque cas d'usage est accompagné d'un tableau indiquant la résolution et la fréquence des trames souhaitées. On commence évidemment par le "streaming" (un des auteurs du RFC travaille chez Netflix.) Pouvoir regarder Le Messie via l'Internet est un objectif crucial. Comme les clients sont variés, et que les conditions de leur accès Internet varient en permanence (surtout s'il y a des liens radio), le codec doit tout le temps s'adapter, et pour cela tenir compte des caractéristiques de la perception humaine. L'encodage est typiquement fait une fois et une seule, lorsque la vidéo est chargée sur les serveurs. Pour cet usage :

- On peut accepter un encodeur complexe et coûteux en ressources, puisque l'encodage sera fait sur de grosses machines, et une seule fois,
- Le décodeur, par contre, doit être simple, car beaucoup de monde devra faire le décodage, et pas forcément sur des machines performantes,
- En entrée, il faut pouvoir accepter beaucoup de formats, et de types de contenus, parfois avec des particularités (le grain d'un film peut être délibéré, l'encodage ne doit pas le supprimer).

Proche du "streaming" mais quand même différent, il y a le partage de vidéos, usage popularisé par YouTube mais également accessible en utilisant du logiciel libre et décentralisé, comme PeerTube. C'est le monde de l'UGC : M. Michu filme des violences policières avec son petit ordiphone (ou avec sa GoPro), et les diffuse au monde entier grâce aux services de partage de vidéos.

Après le "streaming" et le partage de vidéos, la télévision sur IP. Il s'agit de distribuer de la télévision traditionnelle sur IP, pour permettre de se réduire le cerveau en regardant Hanouna. Cet usage se décompose en deux groupes : l'"unicast", où on envoie l'émission à un seul destinataire, c'est la VoD, et le "multicast" où on diffuse à un grand nombre d'utilisateurs simultanément. Ce dernier est sans doute de moins en moins utilisé, à part pour les événements sportifs et les cérémonies. Le RFC se concentre sur la distribution de télévision au-dessus d'un réseau « géré » (comme si les autres ne l'étaient pas...) où on peut garantir une certaine QoS. C'est par exemple le cas d'un FAI qui envoie les chaînes de télévision aux "boxes" des abonnés. Si on diffuse un match de foot, on souhaite que le but qui décide du sort du match atteigne les ordinateurs sans trop de retard sur les hurlements bestiaux de supporters utilisant la diffusion hertzienne, hurlements qu'on entendra dehors. Pour compliquer un peu les choses, on souhaite que l'utilisateur puisse mettre en pause, reprendre, etc.

Et la vidéoconférence ? Là aussi, le délai d'acheminement est crucial, pour qu'on n'ait pas l'impression de parler avec un astronaute perdu sur Mars. (Le RFC suggère 320 millisecondes au grand maximum, et de préférence moins de 100 ms.)

Il y a aussi le partage d'écran, où on choisit de partager l'écran de son ordinateur avec une autre personne. Comme pour la vidéoconférence, il faut du temps réel, encoder au fur et à mesure avec le plus petit retard possible.

Et les jeux vidéo, alors ? Les jeux ont souvent un contenu riche (beaucoup de vidéos). Dans certains cas, le moteur de jeu est quelque part sur l'Internet et, à partir d'un modèle 3D, génère des vidéos qu'il faut envoyer aux joueurs. Là encore, il faut être rapide : si un joueur flingue un zombie, les autres participants au jeu doivent le voir « tout de suite ». (Nombreux détails et exemples, comme GeForce Grid ou Twitch, dans le document N36771 « "Game streaming requirement for Future Video Coding" », de l'ISO/IEC JTC 1/SC 29/WG 11.)

Enfin, le RFC termine la liste des applications avec la vidéosurveillance. Dans le roman « 1984 », George Orwell ne donne aucun détail technique sur le fonctionnement du télécran, et notamment sur

la connexion qui permet au Parti de savoir ce qui se passe chez Winston. Ici, l'idée est de connecter les caméras de vidéosurveillance à l'Internet et de faire remonter les flux vidéo à un central où on les regarde et où on les analyse. Comme l'État peut souhaiter disposer de très nombreuses caméras, le coût matériel est un critère important.

Après cet examen des cas d'usage possibles, la section 4 du RFC est le cahier des charges à proprement parler. D'abord, les exigences générales, qui concernent toutes les applications. Évidemment, le futur codec vidéo de l'Internet doit encoder efficacement, donc, dit le RFC, mieux que les codecs existants, H.265 ou VP9. Évidemment encore, l'interopérabilité entre les différentes mises en œuvre du codec est cruciale, et cela dépend en partie de l'IETF, qui devra écrire une norme claire et cohérente. Cette norme se déclinera en profils, qui limiteront certains aspects du codec, pour des cas d'usages spécifiques.

Le RFC exige également qu'il soit possible de gérer des extensions au format, pour les questions futures, tout en maintenant la compatibilité (un nouveau décodeur doit pouvoir décoder les vieilles vidéos). Le format doit permettre huit bits de couleur (par couleur) au minimum, et de préférence douze, gérer le YCbCr, permettre des résolutions quelconques, autoriser l'accès direct à un point donné de la vidéo, et bien d'autres choses encore. Et le tout doit pouvoir marcher en temps réel (aussi bien l'encodage que le décodage.) Par contre, pour l'encodage de haute qualité, le RFC autorise les encodeurs à être jusqu'à dix fois plus complexes que pour les formats existants : les ressources des ordinateurs croissent plus vite que la capacité des réseaux, et il est donc raisonnable de faire davantage d'efforts pour encoder.

Les réseaux sont imparfaits, et il faut donc gérer les erreurs. Outre les mécanismes présents dans la couche transport, le RFC demande que des mécanismes spécifiques à la vidéo soient également inclus dans le format. En parlant de couche transport, comme le but est de faire un codec vidéo pour l'Internet, le RFC rappelle l'importance de prévoir comment encapsuler le futur format dans les protocoles de transport existants.

Les exigences ci-dessus étaient impératives. Il y en a également des facultatives, par exemple un nombre de bits pour coder chaque couleur allant jusqu'à 16, la gestion du RGB, celle de la transparence. . . Toujours dans les demandes facultatives, le RFC note qu'il serait bon que le format choisi n'empêche pas le parallélisme dans le traitement des vidéos, ce qui implique, par exemple, qu'on ne soit pas obligé de traiter toutes les images précédant une image pour encoder celle-ci. Et qu'on puisse utiliser des algorithmes qui soient acceptables pour les SIMD/GPU.

Ce n'est pas tout de faire une liste au Père Noël avec toutes les jolies choses qu'on voudrait voir dans le nouveau format, encore faut-il une bonne méthodologie pour évaluer les futures propositions. La section 5 décrit les tests à effectuer, avec plusieurs débits différents. Ces tests seront ensuite faits en comparant un codec de référence (HEVC ou VP9) aux codecs candidats pour répondre à ce cahier des charges. Comme le codec utilisera certainement des techniques dépendant de la perception humaine, ces tests « objectifs » ne suffisent pas forcément, et le RFC prévoit également des tests subjectifs, avec la procédure MOS (cf. le document ISO/IEC « PDTR 29170-1 », première partie.)

Et un petit point de sécurité pour finir : le RFC rappelle qu'encodeur et surtout décodeur doivent être paranoïaques, afin d'éviter qu'une vidéo un peu inhabituelle ne mène à des consommations de ressources déraisonnables, voire épuisant la machine.

Pour l'instant, le candidat sérieux pour répondre à ce cahier des charges est le codec AV1, notamment poussé par l'Alliance for Open Media. Mais il ne semble pas tellement se rapprocher de la normalisation formelle. Le groupe de travail ne semble plus très actif.