

Groupe de travail Réseau
Request for Comments : 3551
STD 65
 RFC rendue obsolète : 1890
 Catégorie : Norme

H. Schulzrinne, Columbia University
 S. Casner, Packet Design

juillet 2003
 Traduction Claude Brière de L'Isle

Profil RTP pour conférences audio et vidéo avec contrôle minimal

Statut du présent mémoire

Le présent document spécifie un protocole de normalisation Internet pour la communauté Internet, et appelle à discussion et suggestions en vue de son amélioration. Prière de s'en rapporter à l'édition en cours des "Internet Official Protocol Standards" (normes officielles du protocole Internet) (STD 1) pour connaître l'état de la normalisation et le statut du présent protocole. La distribution du présent mémoire n'est soumise à aucune restriction.

Déclaration de copyright

Copyright (C) The Internet Society (2003). Tous droits réservés

Résumé

Le présent document décrit un profil appelé "RTP/AVP" pour l'usage du protocole de transport en temps réel (RTP), version 2, et du protocole de contrôle associé, RTCP, au sein de conférences audio et vidéo multi participants avec contrôle minimal. Il donne l'interprétation convenable des champs génériques dans la spécification RTP pour les conférences audio et vidéo. En particulier, ce document définit un ensemble de transpositions par défaut de numéros de type de charge utile en codages.

Le présent document décrit aussi comment les données audio et vidéo peuvent être portées au sein de RTP. Il définit un ensemble de codages standard et leurs noms lorsqu'ils sont utilisés dans RTP. Les descriptions donnent des pointeurs sur des mises en œuvre de référence et les normes détaillées. Le présent document est destiné à aider à la mise en œuvre d'applications multimédia audio, vidéo et autres en temps réel.

Le présent mémoire rend obsolète la RFC 1890. Il est pour la plus grande part rétro compatible, sauf pour les fonctions supprimées parce que deux mises en œuvre interopérables n'ont pu être trouvées. Les ajouts à la RFC 1890 codifient les pratiques existantes dans l'utilisation des formats de charge utile sous ce profil et incluent de nouveaux formats de charge utile définis depuis la publication de la RFC 1890.

Table des matières

1. Introduction.....	2
1.1 Terminologie.....	2
2. Format de paquet RTP et RTCP et comportement du protocole.....	2
3. Enregistrement de codages supplémentaires.....	3
4. Audio.....	4
4.1 Règles indépendantes du codage.....	4
4.2 Recommandations pour le fonctionnement.....	6
4.3 Lignes directrices pour les codages audio fondés sur l'échantillonnage.....	6
4.4 Lignes directrices pour les codages audio fondés sur la trame.....	6
4.5 Codages audio.....	7
5. Vidéo.....	17
5.1 CelB.....	17
5.2 JPEG.....	17
5.3 H261.....	18
5.4 H263.....	18
5.5 H263-1998.....	18
5.6 MPV.....	18
5.7 MP2T.....	18
5.8 nv.....	18
6. Définitions des types de charge utile.....	18
7. RTP sur TCP et protocoles similaires de flux d'octets.....	20
8. Allocation d'accès.....	20
9. Changements depuis la RFC 1890.....	20
10. Considérations pour la sécurité.....	22

11. Considérations relatives à l'IANA.....	22
12. Références.....	22
12.1 Références normatives.....	23
12.2 Références pour information.....	23
13. Localisation actuelle des ressources concernées.....	24
14. Remerciements.....	24
15. Déclaration de droits de propriété intellectuelle.....	25
16. Adresse des auteurs.....	25
17. Déclaration complète des droits de reproduction.....	25

1. Introduction

Le présent profil définit les aspects de RTP qui ne sont pas spécifiés dans la définition du protocole RTP version 2 de la RFC 3550 [1]. Ce profil est destiné à être utilisé au sein de conférences audio et vidéo avec contrôle de session minimal. En particulier, aucune prise en charge de la négociation des paramètres ou du contrôle d'adhésion n'est fournie. Le profil est supposé utile dans les sessions où aucune négociation ni contrôle d'adhésion ne sont utilisés (par exemple, utilisant des types de charge utile statiques et les indications d'adhésion fournies par RTCP), mais ce profil peut aussi être utile en conjonction avec un protocole de contrôle de niveau supérieur.

L'usage du présent profil peut être implicite dans l'utilisation des applications appropriées ; il peut n'y avoir pas d'indication explicite du numéro d'accès, de l'identifiant du protocole ou de choses comme cela. Des applications comme des répertoires de session peuvent utiliser le nom spécifié pour ce profil à la Section 11.

D'autres profils peuvent faire des choix différents pour les éléments spécifiés ici.

Le présent document définit aussi un ensemble de codages et de formats de charge utile pour l'audio et la vidéo. Ces descriptions de format de charge utile ne sont incluses ici que dans un souci pratique car elles sont trop petites pour justifier des documents à part. L'utilisation de ces formats de charge utile N'EST PAS EXIGÉE pour le présent profil. Seule la liaison de certains des formats de charge utile aux numéros de type de charge utile statique des tableaux 4 et 5 est normative.

1.1 Terminologie

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" dans ce document sont à interpréter comme décrit dans la RFC 2119 [2] et indiquent les niveaux d'exigence pour les mises en œuvre conformes à ce profil RTP.

Le présent document définit le terme "type de support" comme divisant les codages de contenus audio et vidéo en trois classes : audio, vidéo et audio/vidéo (entrelacé).

2. Format de paquet RTP et RTCP et comportement du protocole

La section "Spécifications des profils et formats de charge utile RTP" de la RFC 3550 énumère un certain nombre d'éléments qui peuvent être spécifiés ou modifiés dans un profil. La présente section traite de ces éléments. Généralement, ce profil suit les aspects par défaut et/ou recommandés de la spécification RTP.

en-tête de données RTP : le format standard de l'en-tête fixe de données RTP (un bit marqueur) est utilisé.

types de charge utile : les types de charge utile statiques sont définis à la Section 6.

ajouts d'en-tête de données RTP : aucun champ additionnel fixe n'est ajouté à l'en-tête de données RTP.

extensions d'en-tête de données RTP : aucune extension d'en-tête RTP n'est définie, mais les applications qui fonctionnent sous le présent profil PEUVENT utiliser de telles extensions. Et donc, les applications NE DEVRAIENT PAS supposer que le bit X de l'en-tête RTP est toujours à zéro et elles DEVRAIENT être prêtes à ignorer l'extension d'en-tête. Si une extension d'en-tête est définie à l'avenir, cette définition DOIT spécifier le contenu des 16 premiers bits d'une façon telle que plusieurs extensions différentes puissent être identifiées.

types de paquet RTCP : aucun type de paquet RTCP supplémentaire n'est défini par la présente spécification de profil.

intervalle de rapports RTCP : Les constantes suggérées sont à utiliser pour le calcul de l'intervalle de rapports RTPC. Les sessions qui fonctionnent sous ce profil PEUVENT spécifier un paramètre distinct pour la bande passante de trafic RTCP plutôt que d'utiliser la fraction par défaut de la bande passante de session. La bande passante du trafic RTCP PEUT être divisée en deux paramètres de session distincts pour les participants qui sont des envoyeurs actifs de données et pour ceux qui ne le sont pas. Suivant la recommandation de la spécification RTP [1] que 1/4 de la bande passante RTCP soit dédiée aux envoyeurs de données, les valeurs par défaut RECOMMANDÉES pour ces deux paramètres seraient respectivement 1,25 % et 3,75 %. Pour une session particulière, la bande passante RTCP pour les non envoyeurs de données PEUT être réglée à zéro lors du fonctionnement sur des liaisons unidirectionnelles ou pour des sessions qui n'exigent pas de retour sur la qualité de réception. La bande passante RTCP pour les envoyeurs de données DEVRAIT rester différente de zéro afin que les envoyeurs de rapports puissent encore les envoyer pour la synchronisation inter supports et pour identifier la source par le CNAME. Les moyens par lesquels sont spécifiés le, ou les deux paramètres de bande passante de session RTCP sortent du domaine d'application du présent mémoire.

extension SR/RR : aucune section d'extension n'est définie pour le paquet SR ou RR RTCP.

utilisation de SDES : les applications PEUVENT utiliser tous les éléments de SDES décrits dans la spécification RTP. Alors que les informations de CNAME DOIVENT être envoyées à tous les intervalles de rapports, d'autres éléments DEVRAIENT seulement être envoyés tous les trois intervalles de rapports, avec NAME envoyé sept fois sur huit au sein de ce créneau et les éléments de SDES restants prenant de façon cyclique le huitième créneau, comme défini au paragraphe 6.2.2 de la spécification RTP. En d'autres termes, NAME est envoyé dans les paquets RTCP 1, 4, 7, 10, 13, 16, 19, alors que, disons EMAIL est utilisé dans le paquet RTCP 22.

sécurité : les service de sécurité RTP par défaut sont aussi par défaut sous ce profil.

transposition de chaîne à clé : aucune transposition n'est spécifiée par ce profil.

encombrement : RTP et ce profil peuvent être utilisés dans le contexte d'un service réseau amélioré, par exemple, de services intégrés (RFC 1633) [4] ou de services différenciés (RFC 2475) [5], ou ils peuvent être utilisés avec un service au mieux.

Si un service amélioré est utilisé, les receveurs RTP DEVRAIENT surveiller la perte de paquets pour s'assurer que le service qui était demandé est réellement fourni. S'il ne l'est pas, ils DEVRAIENT alors supposer qu'ils reçoivent un service au mieux et se conduire en conséquence.

Si le service au mieux est utilisé, les receveurs RTP DEVRAIENT surveiller les pertes de paquet pour s'assurer que le taux de perte de paquet reste dans des limites acceptables. La perte de paquet est considérée comme acceptable si un flux TCP à travers le même chemin de réseau et rencontrant les mêmes conditions de réseau réalisait un débit moyen, mesuré sur une échelle de temps raisonnable, qui ne serait pas inférieur à ce que le flux RTP réalise. Cette condition peut être satisfaite en mettant en œuvre des mécanismes de contrôle d'encombrement pour adapter le taux de transmission (ou le nombre de couches souscrites pour une session de diffusion groupée en couches), ou en s'arrangeant pour qu'un receveur quitte la session si le taux de pertes est inacceptablement élevé.

La comparaison avec TCP ne peut pas être spécifiée exactement, mais elle est prévue comme une comparaison "d'ordre de grandeur" en temps et en débit. Le temps pendant lequel le débit TCP est mesuré est le temps d'aller-retour de la connexion. Par nature, cette exigence établit qu'il n'est pas acceptable de déployer une application (utilisant RTP ou tout autre protocole de transport) sur l'Internet au mieux qui consomme arbitrairement la bande passante et n'est pas en compétition loyale avec TCP dans le même ordre de grandeur.

protocole sous-jacent : le profil spécifie l'utilisation de RTP sur UDP aussi bien que TCP en envoi individuel et en diffusion groupée. (Cela n'empêche pas l'utilisation de ces définitions quand RTP est porté sur d'autres protocoles de couche inférieure.)

transposition de transport : on utilise la transposition standard de RTP et RTCP en adresses de niveau transport.

encapsulation : le présent profil laisse aux applications la spécification de l'encapsulation de RTP dans les protocoles autres que UDP.

3. Enregistrement de codages supplémentaires

Le présent profil fait la liste d'un ensemble de codages, chacun d'eux comportant une compression ou représentation de données de support particulière plus un format de charge utile pour l'encapsulation au sein de RTP. Certains de ces formats de charge utile sont spécifiés ici, alors que d'autres sont spécifiés dans d'autres RFC. Il est prévu que des codages supplémentaires au delà de l'ensemble énuméré ici soient créés à l'avenir et spécifiés dans des RFC de format de charge utile supplémentaires.

Le présent profil alloue aussi à chaque codage un nom abrégé qui PEUT être utilisé par des protocoles de contrôle de niveau supérieur, tels que le protocole de description de session (SDP, *Session Description Protocol*) RFC 2327 [6], pour identifier les codages choisis pour une session RTP particulière.

Dans certains contextes, il peut être utile de se référer à ces codages sous la forme d'un type de contenu MIME. Pour le rendre plus facile, la RFC 3555 [7] donne les enregistrements de tous les noms de codages énumérés ici comme des noms de sous-type MIME sous les types MIME "audio" et "vidéo" à travers la procédure d'enregistrement MIME telle que spécifiée dans la RFC 2048 [8].

Tous les codages supplémentaires spécifiés pour utilisation avec ce profil (ou d'autres) peuvent aussi se voir allouer des noms enregistrés comme sous-types MIME auprès de l'Autorité d'allocation des numéros de l'Internet (IANA). Ce registre donne les moyens de s'assurer que les noms alloués aux codages supplémentaires restent uniques. La RFC 3555 spécifie les informations qui sont requises pour l'enregistrement des codages RTP.

En plus de l'allocation de noms aux codages, ce profil alloue aussi des numéros de type de charge utile statique RTP à certains d'entre eux. Cependant, le numéro de type de charge utile est relativement étroit et ne peut s'accommoder d'allocations pour tous les codages existants et futurs. Durant les premiers stades du développement de RTP, il était nécessaire d'utiliser des types de charge utile alloués de façon statique parce qu'aucun autre mécanisme n'avait été spécifié pour lier le codage et le type de charge utile. Il était prévu que des moyens non RTP sortant du domaine d'application du présent mémoire (comme des protocoles de services d'annuaire ou d'invitation) seraient spécifiés pour établir une transposition dynamique entre un type de charge utile et un codage. Maintenant, des mécanismes de définition de liaisons de type de charge utile dynamiques ont été spécifiés dans le protocole de description de session (SDP) et dans d'autres protocoles tels que la Recommandation UIT-T H.323/H.245. Ces mécanismes associent le nom enregistré du codage/format de charge utile, avec tout paramètre supplémentaire requis, comme le débit d'horloge de l'horodatage RTP et le nombre de canaux, avec un numéro de type de charge utile. Cette association n'est effective que pour la durée de la session RTP dans laquelle est faite la liaison dynamique de type de charge utile. Cette association ne s'applique qu'à la session RTP pour laquelle elle est faite, et donc les numéros peuvent être réutilisés pour des codages différents dans des sessions différentes de sorte que la limitation de l'espace des numéros est évitée.

Le présent profil réserve les numéros de type de charge utile dans la gamme 96-127 exclusivement à l'allocation dynamique. Les applications DEVRAIENT d'abord utiliser les valeurs dans cette gamme pour des types de charge utiles dynamiques. Ces applications qui ont besoin de définir plus de 32 types de charges utiles dynamiques PEUVENT lier des codes en dessous de 96, auquel cas il est RECOMMANDÉ que les numéros de type de charge utile non alloués soient utilisés d'abord. Cependant, les types de charge utile alloués de façon statique sont des liaisons par défaut et PEUVENT être liés de façon dynamique à de nouveaux codages si nécessaire. Redéfinir des types de charge utile en dessous de 96 peut causer un fonctionnement incorrect si on tente de se joindre à une session sans obtenir les informations de description de session qui définissent les types de charge utile dynamiques.

Les types de charge utile dynamiques NE DEVRAIENT PAS être utilisés sans un mécanisme bien défini pour indiquer la transposition. Les systèmes qui s'attendent à interopérer avec d'autres qui fonctionnent sous ce profil NE DEVRAIENT PAS faire leurs propres allocations de codages propriétaires à des types de charge utile fixés particuliers.

La présente spécification établit comme politique de ne pas allouer de types de charge utile statiques supplémentaires au delà de ceux définis dans le présent document. Fixer cette politique évite le problème d'essayer de créer un ensemble de critères d'acceptation des allocations statiques et encourage les mises en œuvre et le déploiement des mécanismes de type de charge utile dynamique.

L'ensemble final des allocations de type de charge utile figure dans les tableaux 4 et 5.

4. Audio

4.1 Règles indépendantes du codage

Comme la capacité de supprimer les silences est une des principales motivations de l'utilisation des paquets pour la transmission de la voix, l'en-tête RTP porte à la fois un numéro de séquence et un horodatage pour permettre au receveur de distinguer les paquets perdus et les périodes où aucune donnée n'est transmise. La transmission discontinue (suppression de

silence) PEUT être utilisée avec tous les formats de charge utile audio. Les receveurs DOIVENT supposer que les envoyeurs peuvent supprimer les silences sauf si c'est interdit par une signalisation qui est spécifiée ailleurs. (Même si l'émetteur ne supprime pas les silences, le receveur devrait être prêt à traiter des périodes où aucune donnée n'est présente car des paquets peuvent être perdus.)

Certains formats de charge utile (voir aux paragraphes 4.5.3 et 4.5.6) définissent une trame de "descripteur d'insertion de silence" ou "bruit de confort" pour spécifier des paramètres de bruit artificiel qui peut être généré durant une période de silence pour approximer le bruit de fond à la source. Pour d'autres formats de charge utile, un bruit de confort (CN, *Comfort Noise*) générique est spécifié dans la RFC 3389 [9]. Lorsque le format de charge utile CN est utilisé avec un autre format de charge utile, les valeurs différentes dans le champ Type de charge utile RTP distinguent les paquets de bruit de confort de ceux du format de charge utile choisi.

Pour les applications qui envoient soit pas de paquet, soit des paquets occasionnels de bruit de confort durant les silences, le premier paquet d'un jet de parole, c'est-à-dire, le premier paquet après une période de silence durant laquelle les paquets n'ont pas été transmis de façon contiguë, DEVRAIT être distingué en réglant le bit marqueur à un dans l'en-tête de données RTP. Le bit marqueur dans tous les autres paquets est à zéro. Le commencement d'un jet de parole PEUT être utilisé pour ajuster le délai de reproduction pour refléter des délais réseau changeants. Les applications sans suppression de silence DOIVENT régler le bit marqueur à zéro.

Le débit d'horloge RTP utilisé pour générer l'horodatage RTP est indépendant du nombre de canaux et du codage ; il est normalement égal au nombre de périodes d'échantillonnage par seconde. Pour des codages sur N canaux, chaque période d'échantillonnage (disons, 1/8 000 de seconde) génère N échantillons. (Cette terminologie est standard, mais quelque peu trompeuse, car le nombre total d'échantillons générés par seconde est alors le taux d'échantillonnage multiplié par le nombre de canaux.)

Si plusieurs canaux audio sont utilisés, les canaux sont numérotés de gauche à droite, en commençant à un. Dans les paquets RTP audio, les informations provenant de canaux de numéro inférieur précèdent ceux des canaux de numéro supérieur.

Quand il y a plus de deux canaux, la convention suivie par le format d'échange audio AIFF-C DEVRAIT être suivie [3], en utilisant la notation suivante, sauf convention contraire spécifiée pour un codage ou format de charge utile particulier :

l (*left*) gauche
 r (*right*) droite
 c centre
 S (*surround*) autour
 F face
 R (*rear*) derrière

canaux	description	canal					
		1	2	3	4	5	6
2	stéréo	l	r				
3		l	r	c			
4		l	c	r	S		
5		l	Fr	Fc	Sl	Sr	
6		l	lc	c	r	rc	S

Note : La RFC 1890 définissait deux conventions pour le rangement de quatre canaux audio. Comme le rangement est indiqué implicitement par le numéro des canaux, c'était ambigu. Dans cette révision, l'ordre décrit comme "quadriphonique" a été éliminé pour lever l'ambiguïté. Ce choix se fonde sur l'observation que le format audio de consommateur quadriphonique n'est pas devenu très populaire, alors que le son enveloppant (*surround*) l'est devenu.

Les échantillons pour tous les canaux qui appartiennent à un seul instant d'échantillonnage DOIVENT être au sein du même paquet. L'entrelaçage des échantillons provenant des différents canaux dépend du codage. Les lignes directrices générales sont données aux paragraphes 4.3 et 4.4.

La fréquence d'échantillonnage DEVRAIT être tirée de l'ensemble : 8 000, 11 025, 16 000, 22 050, 24 000, 32 000, 44 100 et 48 000 Hz. (Les anciens ordinateurs Apple Macintosh avaient un taux d'échantillonnage natif de 22 254,54 Hz, qui peut être converti en 22 050 avec une qualité acceptable en abandonnant 4 échantillons dans une trame de 20 ms.) Cependant, la plupart des codages audio sont définis pour un ensemble plus restreint de fréquences d'échantillonnage. Les receveurs DEVRAIENT être prêts à accepter de l'audio multi canal, mais PEUVENT choisir de ne jouer que sur un seul canal.

4.2 Recommandations pour le fonctionnement

Les recommandations suivantes sont les paramètres de fonctionnement par défaut. Les applications DEVRAIENT être prêtes à traiter d'autres valeurs. Les gammes données sont destinées à fournir des lignes directrices aux auteurs d'applications, en permettant à un ensemble d'applications conformes à ces lignes directrices d'interopérer sans négociation supplémentaire. Ces lignes directrices ne sont pas destinées à restreindre les paramètres de fonctionnement pour les applications qui peuvent négocier un ensemble de paramètres interopérables, par exemple, à travers un protocole de commande de conférence.

Pour l'audio par paquets, l'intervalle de mise en paquet par défaut DEVRAIT avoir une durée de 20 ms ou d'une trame, selon ce qui est le plus long, sauf notation contraire au Tableau 1 (colonne "ms/paquet"). L'intervalle de mise en paquet détermine le délai minimum de bout en bout ; les paquets plus longs introduisent moins de redondance d'en-tête mais de plus forts délais et rendent la perte de paquet moins remarquable. Pour les applications non interactives telles que de lecture ou pour les liaisons qui ont des contraintes sévères de bande passante, un délai de mise en paquet plus élevé PEUT être utilisé. Un receveur DEVRAIT accepter des paquets représentant entre 0 et 200 ms de données audio. (Pour les codages audio en trames, un receveur DEVRAIT accepter des paquets avec un nombre de trames égal à 200 ms divisé par la durée de trame, arrondi à l'entier supérieur.) Cette restriction permet une taille raisonnable de mémoire tampon pour le receveur.

4.3 Lignes directrices pour les codages audio fondés sur l'échantillonnage

Dans les codages fondés sur l'échantillonnage, chaque échantillon audio est représenté par un nombre fixe de bits. Au sein des données audio compressées, les codes des échantillons individuels peuvent s'étendre sur des frontières d'octet. Un paquet RTP audio peut contenir tout nombre d'échantillons audio, selon les contraintes que le nombre de bits par échantillon multiplié par le nombre d'échantillons par paquet donne ou non un nombre entier d'octets. Les codages fractionnaires produisent moins d'un octet par échantillon.

La durée d'un paquet audio est déterminée par le nombre d'échantillons dans le paquet.

Pour les codages fondés sur l'échantillonnage qui produisent un ou plusieurs octets par échantillon, les échantillons provenant de différents canaux échantillonnés au même instant d'échantillonnage DEVRAIENT être mis en paquets dans des octets consécutifs. Par exemple, pour un codage sur deux canaux, la séquence d'octets est (canal gauche, premier échantillon), (canal droit, premier échantillon), (canal gauche, second échantillon), (canal droit, second échantillon), Pour des codages multi octets, les octets DEVRAIENT être transmis dans l'ordre des octets du réseau (c'est-à-dire, l'octet de poids fort en premier).

La mise en paquet des codages fondés sur l'échantillon produisant moins d'un octet par échantillon est spécifique du codage.

L'horodatage RTP reflète l'instant auquel le premier échantillon du paquet été échantillonné, c'est-à-dire, les plus anciennes informations du paquet.

4.4 Lignes directrices pour les codages audio fondés sur la trame

Les codages fondés sur la trame codent un bloc de longueur fixe de données audio dans un autre bloc de données compressées, normalement aussi de longueur fixe. Pour les codages fondés sur la trame, l'expéditeur PEUT choisir de combiner plusieurs trames de cette sorte en un seul paquet RTP. Le receveur peut dire le nombre de trames contenues dans un paquet RTP, si toutes les trames ont la même longueur, en divisant la longueur de charge utile RTP par la taille de trame audio qui est définie au titre du codage. Cela ne fonctionne pas si on transporte des trames de différentes tailles sauf si les tailles de trame sont premières entre elles. Sinon, les trames DOIVENT indiquer leur taille.

Pour les codecs fondés sur la trame, l'ordre des canaux est défini pour tout le bloc. C'est-à-dire que pour deux canaux audio, les échantillons de droite et gauche DEVRAIENT être codés indépendamment, avec la trame codée pour le canal gauche précédant celle pour le canal droit.

Tous les codecs audio orientés trame DEVRAIENT être capables de coder et décoder plusieurs trames consécutives au sein d'un seul paquet. Comme la taille de trame pour le codec orienté trame est donnée, il n'est nul besoin d'utiliser une désignation distincte pour le même codage, mais on utilise des numéros de trame différents pour chaque paquet.

Les paquets RTP DEVRONT contenir un nombre entier de trames, avec les trames insérées selon l'âge au sein d'un paquet, de sorte que la plus ancienne trame (à exécuter en premier) survienne immédiatement après l'en-tête de paquet RTP. L'horodatage RTP reflète l'instant auquel le premier échantillon dans la première trame a été échantillonné, c'est-à-dire, la plus ancienne information du paquet.

4.5 Codages audio

nom du codage	échantillon/trame	bits/échantillon	taux d'échantillonnage	ms/trame	ms/paquet par défaut
DVI4	échantillon	4	variable	20	
G722	échantillon	8	16 000	20	
G723	trame	N/A	8 000	30	30
G726-40	échantillon	5	8 000	20	
G726-32	échantillon	4	8 000	20	
G726-24	échantillon	3	8 000	20	
G726-16	échantillon	2	8 000	20	
G728	trame	N/A	8 000	2.5	20
G729	trame	N/A	8 000	10	20
G729D	trame	N/A	8 000	10	20
G729E	trame	N/A	8 000	10	20
GSM	trame	N/A	8 000	20	20
GSM-EFR	trame	N/A	8 000	20	20
L8	échantillon	8	variable	20	
L16	échantillon	16	variable	20	
LPC	trame	N/A	8 000	20	20
MPA	trame	N/A	variable	variable	
PCMA	échantillon	8	variable	20	
PCMU	échantillon	8	variable	20	
QCELP	trame	N/A	8 000	20	20
VDVI	échantillon	variable	variable	20	

Tableau 1 : Propriétés des codages audio (N/A : non applicable)

Les caractéristiques des codages audio décrites dans ce document sont présentées dans le Tableau 1 ; elles sont énumérées dans l'ordre de leur type de charge utile au Tableau 4. Alors que la plupart des codecs audio ne sont spécifiés que pour un taux d'échantillonnage fixe, certains algorithmes fondés sur l'échantillon (indiqués par une entrée de "variable" dans la colonne "taux d'échantillonnage" du Tableau 1) peuvent être utilisés avec différents taux d'échantillonnage, résultant en différents débits binaires codés. Lorsque ils sont utilisés avec un taux d'échantillonnage autre que celui pour lequel un type de charge utile statique est défini, des moyens non RTP sortant du domaine d'application du présent mémoire DOIVENT être utilisés pour définir un type dynamique de charge utile et DOIVENT indiquer le débit d'horloge d'horodatage RTP choisi, qui est normalement le même que le taux d'échantillonnage pour l'audio.

4.5.1 DVI4

DVI4 utilise un schéma de codage à modulation par impulsions et codage différentiel adaptatif (ADPCM, *adaptive delta pulse code modulation*) qui a été spécifié par l'association du multimédia interactif (IMA, *Interactive Multimedia Association*) comme "type d'onde ADPCM IMA". Cependant, le codage défini ici comme DVI4 diffère sous trois aspects de la spécification IMA :

- o L'en-tête RTP DVI4 contient la valeur prédite plutôt que la valeur du premier échantillon contenu dans l'en-tête de bloc ADPCM de l'IMA.
- o Les blocs ADPCM de l'IMA contiennent un nombre impair d'échantillons, car le premier échantillon d'un bloc est contenu juste dans l'en-tête (non compressé), suivi par un nombre pair d'échantillons compressés. DVI4 a seulement un nombre pair d'échantillons compressés, utilisant le mot "predict" provenant de l'en-tête pour décoder le premier échantillon.
- o Pour DVI4, les échantillons de quatre bits sont mis en paquet avec le premier échantillon des quatre bits de plus fort poids et le second échantillon dans les quatre bits de moindre poids. Dans le codec ADPCM de l'IMA, les échantillons sont mis en paquet dans l'ordre inverse.

Chaque paquet contient un seul bloc DVI. Le présent profil ne définit que la version à 4 bits par échantillon, alors que IMA spécifie aussi un codage à 3 bits par échantillon.

Le mot "header" (*en-tête*) a pour chaque canal la structure suivante :

```
int16    predict;          /* valeur prédite du premier échantillon provenant du bloc précédent (format L16) */
u_int8   index;           /* indice actuel dans le tableau stepsize */
u_int8   reserved;       /* mis à zéro par l'expéditeur, ignoré par le receveur */
```

Chaque octet suivant l'en-tête contient deux échantillons de 4 bits, et donc le nombre d'échantillons par paquet DOIT être pair parce que il n'y a aucun moyen d'indiquer que le dernier octet serait partiellement rempli.

La mise en paquet des échantillons pour des canaux multiples fera l'objet d'études ultérieures.

L'algorithme ADPCM de l'IMA a été décrit dans le document "Pratiques recommandées par l'IMA pour l'amélioration de la compatibilité audio numérique dans les systèmes multimédia (version 3.0)". Cependant, l'Association pour le multimédia interactif a cessé de fonctionner en 1997. Les ressources de copies archivées de ce document et d'un logiciel de mise en œuvre du codage RTP DVI4 sont décrites à la Section 13.

4.5.2 G722

G722 est spécifié dans la Recommandation UIT-T G.722, "Codage audio à 7 kHz sur 64 kbit/s". Le codeur G.722 produit un flux d'octets, dont chacun DOIT être aligné sur l'octet dans un paquet RTP. Le premier bit transmis dans l'octet G.722, qui est le bit de poids fort de l'échantillon de la sous bande la plus élevée, DOIT correspondre au bit de poids fort de l'octet dans le paquet RTP.

Bien que le taux d'échantillonnage réel de l'audio G.722 soit de 16 000 Hz, le débit d'horloge RTP pour le format de charge utile G722 est de 8 000 Hz parce que cette valeur était allouée par erreur dans la RFC 1890 et doit rester inchangée pour la rétro compatibilité. Le débit d'octet ou débit de paire d'échantillons est de 8 000 Hz.

4.5.3 G723

G723 est spécifié dans la Recommandation UIT-T G.723.1, "Codeur de parole à deux débits pour les communications multimédia émises à 5,3 et 6,3 kbit/s". Le codeur G.723.1 à 5,3/6,3 kbit/s a été défini par l'UIT-T comme codec obligatoire pour les applications de terminaux de visiophonie GSTN de la Recommandation UIT-T H.324. L'algorithme a une spécification de virgule flottante dans l'Annexe B à G.723.1, un algorithme de compression de silence dans l'Annexe A et un schéma de codage de canal échelonné pour les applications sans fil dans l'Annexe C.

La présente Recommandation spécifie une représentation codée qui peut être utilisée pour compresser les composants du signal de parole de services multimédia à un très bas débit binaire. L'audio est codé dans des trames de 30 ms, avec un délai supplémentaire de 7,5 ms dû à la pré analyse. Une trame G723 peut être d'une parmi trois tailles : 24 octets (trame de 6,3 kbit/s), 20 octets (trame de 5,3 kbit/s), ou 4 octets. Ces trames de 4 octets sont appelées trames de descripteur d'insertion de silence (SID, *Silence Insertion Descriptor*) et sont utilisées pour spécifier les paramètres du bruit de fond. Il n'y a pas de restriction sur la façon dont les trames à 4, 20, et 24 octets sont entremêlées. Les deux bits de moindre poids du premier octet de la trame déterminent la taille de la trame et le type du codec :

bits	contenu	octets/trame
00	parole à haut débit (6,3 kbit/s)	24
01	parole à bas débit (5,3 kbit/s)	20
10	trame SID	4
11	réservé	

Il est possible de commuter entre les deux débits à toute frontière de trame de 30 ms. Les deux débits (5,3 kbit/s et 6,3 kbit/s) sont une partie obligatoire du codeur et du décodeur. Les receveurs DOIVENT accepter les deux débits de données et DOIVENT accepter les trames SID à moins que des restrictions à ces capacités aient été signalées. L'enregistrement MIME pour G723 dans la RFC 3555 [7] spécifie les paramètres qui PEUVENT être utilisés avec MIME ou SDP pour restreindre à un seul débit de données ou pour interdire l'utilisation des trames SID. Ce codeur a été optimisé pour représenter la parole avec une qualité presque parfaite aux débits ci-dessus en limitant cependant la complexité.

Le paquetage du flux binaire codé en octets et l'ordre de transmission des octets sont spécifiés dans la Recommandation UIT-T G.723.1 et sont les mêmes que ceux produits par la mise en œuvre de référence du code C de G.723. Pour le débit de données de 6,3 kbit/s, ce paquetage est illustré comme suit, avec les bits d'en-tête (HDR) qui sont toujours "0 0" comme le montre la Figure 1 pour indiquer le fonctionnement à 6,3 kbit/s, et le bit Z est toujours mis à zéro. Le diagramme montre le paquetage binaire dans "l'ordre des octets du réseau", aussi appelé ordre gros-boutien. Les bits de chaque mot de 32 bits sont numérotés de 0 à 31, avec le bit de poids fort à gauche et numéroté 0. Les octets de chaque mot sont transmis avec l'octet de poids fort en premier. Les bits de chaque champ de données sont numérotés dans l'ordre de représentation du flux binaire du codage (bit de moindre poids en premier). Les barres verticales indiquent les frontières entre les fragments de champ.


```

|0 0 0 0|1 1 0 0|1 1 0 0 0 0 0 0|0 0 0 0|0 0 0 0|0 0 0 0|0 0 0 0|
|3 2 1 0|1 0 9 8|1 0 9 8 7 6 5 4|3 2 1 0|3 2 1 0|3 2 1 0|3 2 1 0|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Figure 2 : Paquetage binaire de G.723 (5,3 kbit/s)

Le paquetage des trames SID de G.723.1, qui sont indiquées par les bits d'en-tête (HDR) qui ont le schéma "1 0", est décrit à la Figure 3.

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+-----+-----+-----+-----+-----+-----+-----+
| LPC | HDR | LPC | LPC | GAIN | LPC | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
|0 0 0 0 0 0|1 0|1 1 1 1 0 0 0 0|2 2 1 1 1 1 1 1|0 0 0 0 0 0|2 2|
|5 4 3 2 1 0| | |3 2 1 0 9 8 7 6|1 0 9 8 7 6 5 4|5 4 3 2 1 0|3 2|
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Figure 3 : Paquetage binaire de G.723 en mode SID

4.5.4 G726-40, G726-32, G726-24 et G726-16

La Recommandation UIT-T G.726 décrit, entre autres, l'algorithme recommandé pour la conversion d'un seul canal MIC loi A ou loi μ à 64 kbit/s codé à 8 000 échantillons/s en et de canal à 40, 32, 24, ou 16 kbit/s. La conversion est appliquée au flux MIC en utilisant une technique de transcodage MICDA, modulation par impulsions et codage différentiel adaptatif (*ADPCM, Adaptive Differential Pulse Code Modulation*). La représentation MICDA consiste en une série de codets avec une correspondance biunivoque aux échantillons du flux MIC. Les débits de données G726 de 40, 32, 24 et 16 kbit/s ont des codets de respectivement 5, 4, 3 et 2 bits.

Les codages à 16 et 24 kbit/s ne fournissent pas la qualité de parole parfaite. Ils sont conçus pour utilisation sur des EMCN, équipements de multiplication de circuit numérique (*DCME, Digital Circuit Multiplication Equipment*) surchargés. La Recommandation UIT-T G.726 recommande que les codages à 16 et 24 kbit/s soient alternés avec des codages à plus hauts débits afin de fournir une taille d'échantillon moyenne entre 3,5 et 3,7 bits par échantillon.

Les codages de G.726 sont notés ici G726-40, G726-32, G726-24 et G726-16. Avant 1990, G721 décrivait le codage MICDA à 32 kbit/s et G723 décrivait les codages à 40, 32 et 16 kbit/s. Et donc, G726-32 désigne le même algorithme que G721 dans la RFC 1890.

Un flux de codets G726 ne contient aucune information sur le codage utilisé, et donc, les transitions entre types de codage G726 ne sont pas permis au sein d'une séquence de codets empaquetés. Les applications DOIVENT déterminer le type de codage des codets empaquetés à partir de l'identifiant de charge utile RTP.

Aucune information d'en-tête spécifique de la charge utile ne DEVRA être incluse au titre des données audio. Un flux de codets G726 DOIT être empaqueté en octets comme suit : le premier codet est placé dans le premier octet de telle sorte que le bit de moindre poids du codet s'aligne sur le bit de moindre poids de l'octet, que le second codet soit alors empaqueté de telle sorte que le bit de moindre poids coïncide avec le bit de moindre poids inoccupé dans l'octet. Lorsque un codet complet ne peut être placé dans un octet, les bits qui chevauchent la frontière de l'octet sont placés dans les bits de moindre poids du prochain octet. L'empaquetage DOIT se terminer sur un octet final complètement empaqueté. Le nombre de codets empaquetés va donc être un multiple de 8, 2, 8 et 4 pour, respectivement, G726-40, G726-32, G726-24 et G726-16. Un exemple de schéma d'empaquetage pour des codets G726-32 est donné ci-dessous, où le bit 7 est le bit de moindre poids du premier octet, et le bit A3 est le bit de moindre poids du premier codet :

```

0 1
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-----+-----+-----+-----+-----+
|B B B B|A A A A|D D D D|C C C C| ...
|0 1 2 3|0 1 2 3|0 1 2 3|0 1 2 3|
+-----+-----+-----+-----+

```

Un exemple de schéma d'empaquetage pour des codets G726-24 est donné ci-dessous, avec encore le bit 7 comme bit de moindre poids du premier octet, et le bit A2 comme bit de moindre poids du premier codet :

```

      0                1                2
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
      +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
      |C C|B B B|A A A|F|E E E|D D D|C|H H H|G G G|F F| ...
      |1 2|0 1 2|0 1 2|2|0 1 2|0 1 2|0|0 1 2|0 1 2|0 1 2|0 1 2|
      +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Noter que la direction "petite boutienne" dans laquelle les échantillons sont empaquetés en octets dans les formats de charge utile G726-16, -24, -32 et -40 spécifiés ici est cohérente avec la Recommandation UIT-T X.420, mais est à l'inverse de ce que spécifie l'Annexe E de la Recommandation UIT-T I.366.2 pour le transport AAL2 en ATM. Un second ensemble de formats de charge utile RTP correspondant à la mise en paquet de I.366.2 Annexe E et identifié par les sous-types MIME AAL2-G726-16, -24, -32 et -40 sera spécifié dans un autre document.

4.5.5 G728

G728 est spécifié dans la Recommandation UIT-T G.728, "Codage de la parole à 16 kbit/s utilisant la prédiction linéaire à faible délai avec excitation par code".

Un codeur G.728 traduit cinq échantillons audio consécutifs en un indice de codets de 10 bits, résultant en un débit binaire de 16 kbit/s pour l'audio échantillonné à 8 000 échantillons par seconde. Le groupe de cinq échantillons consécutifs est appelé un vecteur. Quatre vecteurs consécutifs, marqués V1 à V4 (où V1 est à exécuter en premier par le receveur) construisent une trame G.728. Les quatre vecteurs de 40 bits sont empaquetés dans cinq octets, marqués B1 à B5. B1 DEVRA être placé en premier dans le paquet RTP.

En se référant à la figure ci-dessous, le principe pour l'ordre des bits est la "conservation de la pondération binaire". Les bits provenant d'un vecteur plus ancien sont de plus fort poids que les bits provenant des vecteurs plus récents. Le MSB de la trame va au MSB de B1 et le LSB de la trame va au LSB de B5.

```

      1          2          3          3
      0          0          0          0          9
      +-----+-----+-----+-----+-----+
      <---V1---><---V2---><---V3---><---V4---> vecteurs
      <--B1--><--B2--><--B3--><--B4--><--B5--> octets
      <----- trame 1 ----->

```

En particulier, B1 contient les huit bits de plus forts poids de V1, le MSB de V1 étant le MSB de B1. B2 contient les deux bits de moindre poids de V1, celui de plus fort poids des deux dans son MSB, et les six bits de plus fort poids de V2. B1 DEVRA être placé en premier dans le paquet RTP et B5 en dernier.

4.5.6 G729

G729 est spécifié dans la Recommandation UIT-T G.729, "Codage de la parole à 8 kbit/s en prédiction linéaire à excitation par séquences codées à structure algébrique conjuguée (CS-ACELP, *conjugate structure-algebraic code excited linear prediction*)". Une version à complexité réduite de l'algorithme G.729 est spécifié à l'Annexe A de la Recommandation UIT-T G.729. Les algorithmes de codage de la parole dans le corps de G.729 et dans l'Annexe A de G.729 sont pleinement interoperables entre eux, de sorte qu'il n'est pas besoin de distinguer entre eux. Une mise en œuvre qui signale ou accepte l'utilisation du format de charge utile G729 peut mettre en œuvre G.729 ou G.729A sauf interdit par de la signalisation supplémentaires spécifiée ailleurs se rapportant spécifiquement au codage plutôt qu'au format de charge utile. Les codecs G.729 et G.729 Annexe A ont été optimisés pour représenter la parole avec une grande qualité, mais G.729 Annexe A troque un peu de qualité vocale pour une réduction d'environ 50 % de la complexité [10]. Voir au paragraphe suivant (4.5.7) les autres débits de données ajoutés dans les annexes ultérieures de G.729. Pour tous les débits de données, la fréquence d'échantillonnage (et le débit d'horloge d'horodatage RTP) est de 8 000 Hz.

Un algorithme de détecteur d'activité vocale (VAD) et de générateur de bruit de confort (CNG) dans l'Annexe B de G.729 est RECOMMANDÉ pour les applications de voix et données numériques simultanées et peut être utilisé conjointement avec G.729 ou G.729 Annexe A. Une trame G.729 ou G.729 Annexe A contient 10 octets, alors que la trame de bruit de confort de G.729 Annexe B occupe 2 octets. Les receveurs DOIVENT accepter les trames de bruit de confort si l'interdiction de leur utilisation n'a pas été signalée. L'enregistrement MIME pour G729 dans la RFC 3555 [7] spécifie un paramètre qui PEUT être utilisé avec MIME ou SDP pour interdire l'usage de trames de bruit de confort.

Un paquet RTP G729 peut consister en zéro, une ou plusieurs trames G.729 ou G.729 Annexe A, suivi par zéro ou une trame G.729 Annexe B. La présence d'une trame de bruit de confort peut se déduire de la longueur de la charge utile RTP.

L'intervalle de mise en paquet par défaut est de 20 ms (deux trames), mais dans certaines situations il peut être souhaitable d'envoyer des paquets à 10 ms. Un exemple serait une transition de la parole au bruit de confort dans le premier paquet à 10 ms. Pour certaines applications, un intervalle de mise en paquet plus long peut être nécessaire pour réduire le débit de paquets.

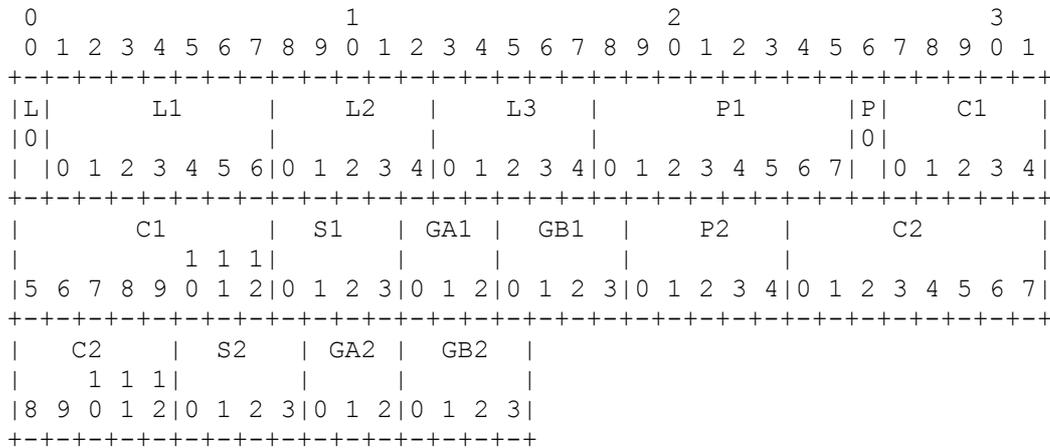


Figure 4 : Paquetage binaire de G.729 et G.729A

Les paramètres transmis d'une trame G.729/G.729A à 10 ms, consistant en 80 bits, sont définis dans la Recommandation G.729, Tableau 8/G.729. La transposition de ces paramètres est donnée ci-dessous à la Figure 4. Le diagramme montre le paquetage binaire dans "l'ordre des octets du réseau", aussi appelé ordre gros-boutien. Les bits de chaque mot de 32 bits sont numérotés de 0 à 31, le bit de poids fort à gauche et numéroté 0. Les octets de chaque mot sont transmis le plus fort poids en premier. Les bits de chaque champ de données sont numérotés dans l'ordre où ils sont produits par la mise en œuvre de référence de code C de G.729.

Le paquetage de la trame de bruit de confort de G.729 Annexe B est indiqué ci-dessous à la Figure 5.

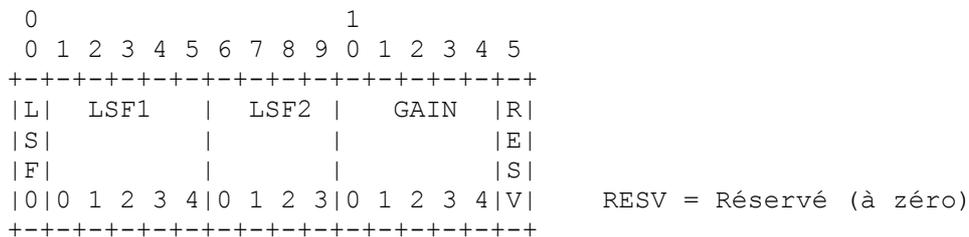


Figure 5 : Paquetage binaire de G.729 Annexe B

4.5.7 G729D et G729E

Les Annexes D et E à la Recommandation UIT-T G.729 donnent des débits de données supplémentaires. Comme le débit des données n'est pas signalé dans le flux binaire, les différents débits de données reçoivent des nom de codage RTP distincts qui sont transposés en numéros de type de charge utile distincts. G729D indique un mode de codage à 6,4 kbit/s (G.729 Annexe D, pour une réduction momentanée de la capacité du canal), tandis que G729E indique un mode à 11,8 kbit/s (G.729 Annexe E, pour des performances améliorées avec une large gamme de signaux d'entrée à bas débit, par exemple, de la musique et du bruit de fond). L'Annexe E a deux modes de fonctionnement, adaptif différé et adaptif anticipé, qui sont signalés par les deux premiers bits de chaque trame (les deux bits de plus fort poids du premier octet).

L'algorithme de détecteur d'activité vocale (VAD) et de générateur de bruit de confort (CNG) spécifié dans l'Annexe B de G.729 peut être utilisé avec les trames de l'Annexe D et de l'Annexe E en plus des trames de G.729 et G.729 Annexe A. Les détails de l'algorithme pour le fonctionnement des Annexes D et E avec la CNG de l'Annexe B sont spécifiés dans les Annexes F et G de G.729. Noter que les Annexes F et G n'introduisent aucun nouveau codage. Les receveurs DOIVENT accepter les trames de bruit de confort si l'interdiction de leur usage n'a pas été signalée. L'enregistrement MIME pour G729D et G729E dans la RFC 3555 [7] spécifie un paramètre qui PEUT être utilisé avec MIME ou SDP pour interdire l'usage des trames de bruit de confort.

Pour G729D, un paquet RTP peut comporter zéro, une ou plusieurs trames G.729 Annexe D, suivies par zéro ou une trame G.729 Annexe B. De même, pour G729E, un paquet RTP peut comporter zéro, une ou plusieurs trames G.729 Annexe E,

suivies par zéro ou une trame G.729 Annexe B. La présence d'une trame de bruit de confort peut être déduite de la longueur de la charge utile RTP.

Un seul paquet RTP doit contenir des trames d'un seul débit de données, facultativement suivies par une trame de bruit de confort. Le débit de données peut être changé d'un paquet à l'autre en changeant le numéro de type de charge utile. Les Annexes D, E et H de G.729 décrivent ce que doivent être les algorithmes de codage et de décodage pour s'accommoder d'un changement du débit des données.

Pour G729D, les bits d'une trame G.729 Annexe D sont formatées comme indiqué ci-dessous à la Figure 6 (cf. Tableau D.1/G.729). La longueur de la trame est de 64 bits.

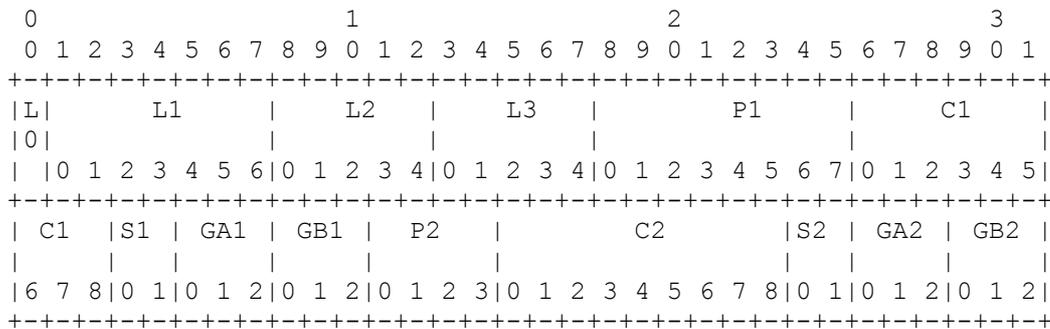


Figure 6 : Paquetage binaire de G.729 Annexe D

Le débit binaire net pour l'algorithme G.729 Annexe E est de 11,8 kbit/s et un total de 118 bits est utilisé. Deux bits sont ajoutés comme bits "à ne pas prendre en compte" pour compléter un nombre entier d'octets pour la trame. Pour G729E, les bits d'une trame de données sont formatés comme indiqué dans les deux diagrammes suivants (cf. Tableau E.1/G.729). Les champs pour le mode adaptatif anticipé de G729E sont empaquetés comme indiqué à la Figure 7.

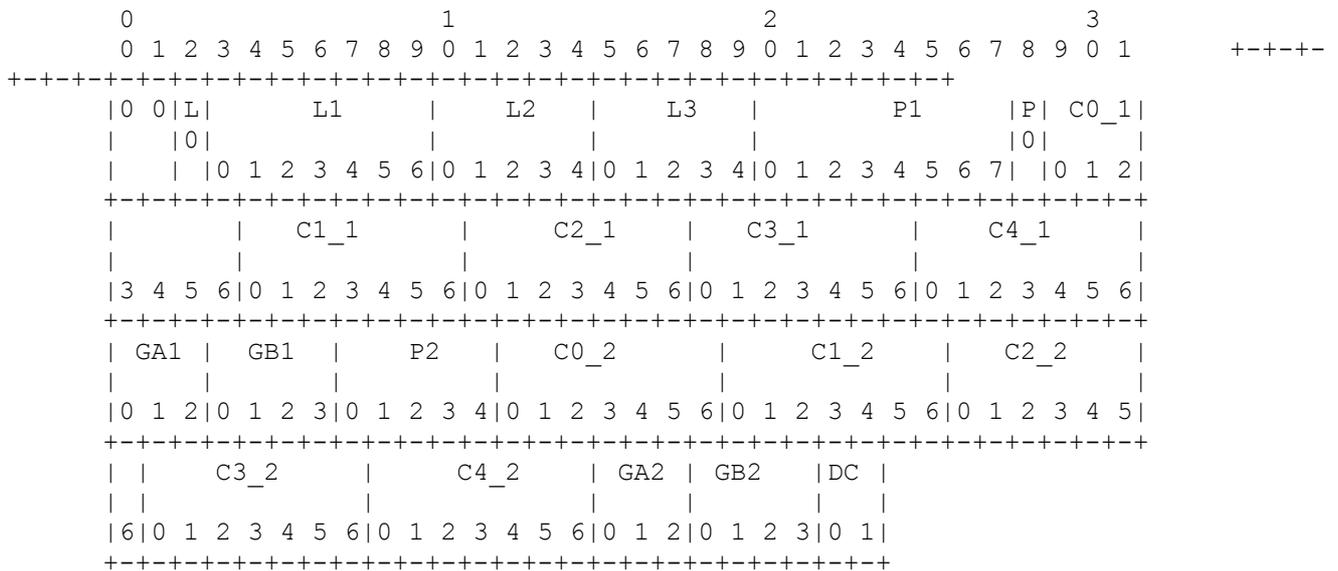


Figure 7 : Paquetage binaire de G.729 Annexe E (mode d'adaptation vers l'avant)

Les champs pour le mode adaptatif différé de G729E sont empaquetés comme indiqué à la Figure . 8.

Octet	Bit 0	Bit 1	Bit 2	Bit 3	Bit 4	Bit 5	Bit 6	Bit 7
0	1	1	0	1	LARc0.0	LARc0.1	LARc0.2	LARc0.3
1	LARc0.4	LARc0.5	LARc1.0	LARc1.1	LARc1.2	LARc1.3	LARc1.4	LARc1.5
2	LARc2.0	LARc2.1	LARc2.2	LARc2.3	LARc2.4	LARc3.0	LARc3.1	LARc3.2
3	LARc3.3	LARc3.4	LARc4.0	LARc4.1	LARc4.2	LARc4.3	LARc5.0	LARc5.1
4	LARc5.2	LARc5.3	LARc6.0	LARc6.1	LARc6.2	LARc7.0	LARc7.1	LARc7.2
5	Nc0.0	Nc0.1	Nc0.2	Nc0.3	Nc0.4	Nc0.5	Nc0.6	bc0.0
6	bc0.1	Mc0.0	Mc0.1	xmaxc00	xmaxc01	xmaxc02	xmaxc03	xmaxc04
7	xmaxc05	xmc0.0	xmc0.1	xmc0.2	xmc1.0	xmc1.1	xmc1.2	xmc2.0
8	xmc2.1	xmc2.2	xmc3.0	xmc3.1	xmc3.2	xmc4.0	xmc4.1	xmc4.2
9	xmc5.0	xmc5.1	xmc5.2	xmc6.0	xmc6.1	xmc6.2	xmc7.0	xmc7.1
10	xmc7.2	xmc8.0	xmc8.1	xmc8.2	xmc9.0	xmc9.1	xmc9.2	xmc10.0
11	xmc10.1	xmc10.2	xmc11.0	xmc11.1	xmc11.2	xmc12.0	xmc12.1	xcm12.2
12	Nc1.0	Nc1.1	Nc1.2	Nc1.3	Nc1.4	Nc1.5	Nc1.6	bc1.0
13	bc1.1	Mc1.0	Mc1.1	xmaxc10	xmaxc11	xmaxc12	xmaxc13	xmaxc14
14	xmax15	xmc13.0	xmc13.1	xmc13.2	xmc14.0	xmc14.1	xmc14.2	xmc15.0
15	xmc15.1	xmc15.2	xmc16.0	xmc16.1	xmc16.2	xmc17.0	xmc17.1	xmc17.2
16	xmc18.0	xmc18.1	xmc18.2	xmc19.0	xmc19.1	xmc19.2	xmc20.0	xmc20.1
17	xmc20.2	xmc21.0	xmc21.1	xmc21.2	xmc22.0	xmc22.1	xmc22.2	xmc23.0
18	xmc23.1	xmc23.2	xmc24.0	xmc24.1	xmc24.2	xmc25.0	xmc25.1	xmc25.2
19	Nc2.0	Nc2.1	Nc2.2	Nc2.3	Nc2.4	Nc2.5	Nc2.6	bc2.0
20	bc2.1	Mc2.0	Mc2.1	xmaxc20	xmaxc21	xmaxc22	xmaxc23	xmaxc24
21	xmaxc25	xmc26.0	xmc26.1	xmc26.2	xmc27.0	xmc27.1	xmc27.2	xmc28.0
22	xmc28.1	xmc28.2	xmc29.0	xmc29.1	xmc29.2	xmc30.0	xmc30.1	xmc30.2
23	xmc31.0	xmc31.1	xmc31.2	xmc32.0	xmc32.1	xmc32.2	xmc33.0	xmc33.1
24	xmc33.2	xmc34.0	xmc34.1	xmc34.2	xmc35.0	xmc35.1	xmc35.2	xmc36.0
25	Xmc36.1	xmc36.2	xmc37.0	xmc37.1	xmc37.2	xmc38.0	xmc38.1	xmc38.2
26	Nc3.0	Nc3.1	Nc3.2	Nc3.3	Nc3.4	Nc3.5	Nc3.6	bc3.0
27	bc3.1	Mc3.0	Mc3.1	xmaxc30	xmaxc31	xmaxc32	xmaxc33	xmaxc34
28	xmaxc35	xmc39.0	xmc39.1	xmc39.2	xmc40.0	xmc40.1	xmc40.2	xmc41.0
29	xmc41.1	xmc41.2	xmc42.0	xmc42.1	xmc42.2	xmc43.0	xmc43.1	xmc43.2
30	xmc44.0	xmc44.1	xmc44.2	xmc45.0	xmc45.1	xmc45.2	xmc46.0	xmc46.1
31	xmc46.2	xmc47.0	xmc47.1	xmc47.2	xmc48.0	xmc48.1	xmc48.2	xmc49.0
32	xmc49.1	xmc49.2	xmc50.0	xmc50.1	xmc50.2	xmc51.0	xmc51.1	xmc51.2

Tableau 3 : Format de charge utile GSM

Dans le paquetage GSM utilisé par RTP, les bits DEVRONT être mis en paquet en commençant par le bit de poids fort. Tous les 160 échantillons, la trame GSM est codée dans une mémoire tampon de 33 octets (264 bits). Chacune de ces mémoires tampon commence par une signature de 4 bits (0xD), suivie par le codage MSB des champs de la trame. Le premier octet contient donc 1101 dans les 4 bits de plus fort poids (0 à 3) et les 4 bits de plus fort poids de F1 (0 à 3) dans les 4 bits de moindre poids (4 à 7). Le second octet contient les 2 bits de moindre poids de F1 dans les bits 0 et 1, et F2 dans les bits 2 à 7, et ainsi de suite. L'ordre des champs dans la trame est décrit au Tableau 2.

4.5.8.2 Noms et numéros de variable GSM

Dans le codage RTP, nous avons le schéma binaire décrit au Tableau 3, où F.i signifie le $i^{\text{ème}}$ bit du champ F, le bit 0 est le bit de plus fort poids, et les bits de chaque octet sont numérotés de 0 à 7 du plus fort au moindre poids.

4.5.9 GSM-EFR

GSM-EFR désigne le codage vocal à plein débit amélioré GSM 06.60, spécifié dans l'ETS 300 726 qui est disponible auprès d'ETSI à l'adresse donnée au paragraphe 4.5.8. Ce codec a une longueur de trame de 244 bits. Pour la transmission en RTP, chaque trame du codec est empaquetée dans une mémoire tampon de 31 octet (248 bit) commençant par une signature de 4 bits 0xC de la même façon que celle spécifiée ici pour le codec original GSM 06.10. L'empaquetage est spécifié dans la spécification technique ETSI TS 101 318.

4.5.10 L8

L8 désigne les échantillons de données audio linéaires, qui utilisent 8 bits de précision avec un décalage de 128, c'est-à-dire que le signal le plus négatif est codé comme zéro.

4.5.11 L16

L16 désigne les échantillons de données audio non compressés, en utilisant une représentation de 16 bits signée avec 65 535 étapes également divisées entre niveau de signal minimum et maximum, allant de -32 768 à 32 767. La valeur est représentée en notation de complément à deux et transmise dans l'ordre des octets du réseau (octet de poids fort en premier).

L'enregistrement MIME pour L16 dans la RFC 3555 [7] spécifie les paramètres qui PEUVENT être utilisés avec MIME ou SDP pour indiquer que la pré amplification analogique a été appliquée au signal avant sa quantification ou pour indiquer qu'un flux audio multi canaux suit une convention de rangement des canaux différente de celle qui est spécifiée au paragraphe 4.1.

4.5.12 LPC

LPC désigne un codage prédictif linéaire expérimental proposé par Ron Frederick, qui se fonde sur une mise en œuvre écrite par Ron Zuckerman envoyée au groupe Usenet comp.dsp le 26 juin 1992. Le codec génère 14 octets pour chaque trame. La taille de trame est réglée à 20 ms, résultant en un débit binaire de 5 600 bit/s.

4.5.13 MPA

MPA désigne de l'audio MPEG-1 ou MPEG-2 encapsulé comme flux élémentaires. Le codage est défini dans les normes ISO/CEI 11172-3 et 13818-3. L'encapsulation est spécifiée dans la RFC 2250 [14].

Le codage peut avoir un des trois niveaux de complexité suivants, appelés Couche I, II et III. La couche choisie ainsi que le taux d'échantillonnage et le nombre de canaux sont indiqués dans la charge utile. Le débit d'horloge de l'horodatage RTP est toujours de 90 000, indépendamment du taux d'échantillonnage. L'audio MPEG-1 prend en charge des taux d'échantillonnage de 32, 44,1, et 48 kHz (ISO/CEI 11172-3, section 1.1 "Domaine d'application"). MPEG-2 prend en charge des taux d'échantillonnage de 16, 22,05 et 24 kHz. Le nombre d'échantillons par trame est fixe, mais la taille des trames va varier selon le taux d'échantillonnage et le débit binaire.

L'enregistrement MIME pour MPA dans la RFC 3555 [7] spécifie des paramètres qui PEUVENT être utilisés avec MIME ou SDP pour interdire le choix d'une couche, du compte de canaux, du taux d'échantillonnage, et du débit binaire.

4.5.14 MICA et MICU

Le MICA et le MICU sont spécifiés dans la Recommandation UIT-T G.711. Les données audio sont codées avec huit bits par échantillon, après échelonnement logarithmique. Le MICU désigne l'échelonnement selon la loi μ , le MICA l'échelonnement selon la loi A. Une description détaillée est donnée par Jayant et Noll [15]. Chaque octet G.711 DEVRA être aligné sur l'octet dans un paquet RTP. Le bit de signe de chaque octet G.711 DEVRA correspondre au bit de plus fort poids de l'octet dans le paquet RTP (c'est-à-dire, en supposant que les échantillons G.711 sont traités comme des octets sur la machine hôte, le bit de signe DEVRA être le bit de plus fort poids de l'octet comme défini par le format de la machine hôte). Les modes 56 kbit/s et 48 kbit/s de G.711 ne sont pas applicables à RTP, car le MICA et le MICU DOIVENT toujours être transmis comme des échantillons à 8 bits.

Voir au paragraphe 4.1 au sujet de la suppression de silence.

4.5.15 QCELP

La norme IS-733, "TR45 : Option de service vocal à haut débit pour systèmes de communications large bande à étalement de spectre", de l'Association des industries électroniques (EIA) et de l'Association des industries de télécommunications (TIA) définit l'algorithme de compression audio QCELP pour une utilisation dans les applications AMRC sans fil. Le codec QCELP compresse chaque 20 millisecondes d'entrées de parole à 8 000 Hz, échantillonnée sur 16 bits dans une des quatre différentes tailles de trame de sortie : taux 1 (266 bits), taux 1/2 (124 bits), taux 1/4 (54 bits) ou taux 1/8 (20 bits). Pour les schémas de parole normaux, il en résulte une sortie moyenne de 6,8 kbit/s pour le mode normal et de 4,7 kbit/s pour le mode à taux réduit. La mise en paquet du codec audio QCELP est décrite dans [16].

4.5.16 RED

Le format de charge utile audio redondant "RED" est spécifié par la RFC 2198 [17]. Il définit le moyen par lequel plusieurs copies redondantes d'un paquet audio peuvent être transmises dans un seul flux RTP. Chaque paquet d'un tel flux contient, en plus des données audio pour cet intervalle de mise en paquet, une copie (plus lourdement compressée) des données d'un intervalle de mise en paquet précédent. Cela permet une approximation de récupération des données provenant des paquets

perdus lors du décodage du paquet suivant, donnant une qualité sonore améliorée comparée à la substitution au silence des paquets perdus.

4.5.17 VDVI

VDVI est une version à débit variable de DVI4, qui donne des débits binaires de parole entre 10 et 25 kbit/s. Elle est spécifiée seulement pour le fonctionnement sur un seul canal. Les échantillons sont mis en paquets dans les octets en commençant par le bit de poids fort. Le dernier octet est empaqueté avec des bits 1 si le dernier échantillon ne remplit pas le dernier octet. Ce bourrage est distinct de celui des codets valides. Le receveur doit détecter le bourrage parce qu'il n'y a pas de compte explicite des échantillons dans le paquet.

Elle utilise le codage suivant :

Codet DVI4	Schéma binaire VDVI
0	00
1	010
2	1100
3	11100
4	111100
5	1111100
6	11111100
7	11111110
8	10
9	011
10	1101
11	11101
12	111101
13	1111101
14	11111101
15	11111111

5. Vidéo

Les paragraphes qui suivent décrivent les codages vidéo qui sont définis dans le présent mémoire et donnent les noms abrégés utilisés pour leur identification. Ces codages vidéo et leurs types de charge utile sont énumérés au Tableau 5.

Tous ces codages vidéo utilisent une fréquence d'horodatage RTP de 90 000 Hz, la même que la fréquence d'horodatage de présentation MPEG. Cette fréquence donne des incréments d'entiers exacts d'horodatage pour les débits de trames normaux de 24 Hz (HDTV) 25 Hz (PAL) 29,97 Hz (NTSC) et 30 Hz (HDTV) et des débits de champ de 50, 59,94 et 60 Hz. Bien que 90 kHz soit le débit RECOMMANDÉ pour les futurs codages vidéo utilisés au sein de ce profil, d'autres débits PEUVENT être utilisés. Cependant, il n'est pas suffisant d'utiliser le débit de trame vidéo (normalement entre 15 et 30 Hz) parce que cela ne fournit pas une résolution adéquate pour les exigences normales de synchronisation lors du calcul de l'horodatage RTP correspondant à l'horodatage NTP dans un paquet SR de RTCP. La résolution d'horodatage DOIT aussi être suffisante pour l'estimation de gigue contenue dans les rapports de receveur.

Pour la plupart de ces codages vidéo, l'horodatage RTP code l'instant d'échantillonnage de l'image vidéo contenue dans le paquet de données RTP. Si une image vidéo occupe plus d'un paquet, l'horodatage est le même sur tous ces paquets. Les paquets provenant de différentes images vidéo sont distinguées par leurs horodatages différents.

La plupart de ces codages vidéo spécifient aussi que le bit marqueur de l'en-tête RTP DEVRAIT être mis à un dans le dernier paquet d'une trame vidéo, et autrement mis à zéro. Et donc, il n'est pas nécessaire d'attendre un paquet suivant avec un horodatage différent pour détecter qu'une nouvelle trame devrait être affichée.

5.1 CeIB

Le codage CELL-B est un codage breveté proposé par Sun Microsystems. Le format de flux d'octets est décrit dans la RFC 2029 [18].

5.2 JPEG

Le codage JËG est spëcifië par les normes ISO 10918-1 et 10918-2. Le format de charge utile RTP est spëcifië dans la RFC 2435 [19].

5.3 H261

Le codage est spëcifië dans la Recommandation UIT-T H.261, "Vidëo codec pour services audiovisuels à p x 64 kbit/s". La mise en paquet et les propriëtës spëcifiques de RTP sont dëcrites dans la RFC 2032 [20].

5.4 H263

Le codage H263 est spëcifië dans la version 1996 de la Recommandation UIT-T H.263, "Codage vidëo pour communication à faible dëbit binaire". La mise en paquet et les propriëtës spëcifiques de RTP sont dëcrites dans la RFC 2190 [21]. Le format de charge utile H263-1998 est RECOMMANDË sur ce profil avec les nouvelles mises en œuvre.

5.5 H263-1998

Ce codage est spëcifië dans la version 1998 de la Recommandation UIT-T H.263, "Codage Vidëo pour communication à faible dëbit binaire". La mise en paquet et les propriëtës spëcifiques de RTP sont dëcrites dans la RFC 2429 [22]. Parce que la version 1998 de H.263 est un surensemble de la syntaxe de 1996, ce format de charge utile peut aussi être utilisé avec la version de 1996 de H.263, et son utilisation est RECOMMANDËE pour les nouvelles mises en œuvre. Ce format de charge utile ne remplace pas la RFC 2190, qui continue d'être utilisée par les mises en œuvre existantes, et peut être requise pour la rëtro compatibilitë dans les nouvelles mises en œuvre. Les mises en œuvre qui utilisent les nouvelles caractëristiques de la version 1998 de H.263 DOIVENT utiliser le format de charge utile dëcrit dans la RFC 2429.

5.6 MPV

MPV dësigne l'utilisation des flux élémentaires de codage vidëo MPEG-1 et MPEG-2 comme spëcifië dans les normes ISO/CEI 11172 et 13818-2, respectivement. Le format de charge utile RTP est spëcifië dans la RFC 2250 [14], Section 3.

L'enregistrement MIME pour MPV dans la RFC 3555 [7] spëcifie un paramètre qui PEUT être utilisé avec MIME ou SDP pour interdire la sëlëction du type vidëo MPEG.

5.7 MP2T

MP2T dësigne l'utilisation de flux de transport MPEG-2, pour l'audio et/ou la vidëo. Le format de charge utile RTP est dëcrit dans la RFC 2250 [14], Section 2.

5.8 nv

Ce codage est mis en œuvre dans le programme 'nv', version 4, dëveloppë chez Xerox PARC par Ron Frederick. Des informations complëmëntaires peuvent être obtenues de l'auteur :

Ron Frederick
Blue Coat Systems Inc.
650 Almanor Avenue
Sunnyvale, CA 94085
USA
mël : ronf@bluecoat.com

6. Dëfinitions des types de charge utile

Les Tableaux 4 et 5 dëfinissent les valeurs de type de charge utile statique de ce profil pour le champ PT de l'en-tête de donnëes RTP. De plus, les valeurs de type de charge utile dans la gamme 96 à 127 PEUVENT être dëfinies de façon dynamique grâce à un protocole de contrôle de confërence, qui sort du domaine d'application du prësnt document. Par exemple, un rëpertoire de session pourrait spëcifier que pour une session donnëe, le type de charge utile 96 indique le codage MICU, un taux d'ëchantillonnage de 8 000 Hz, 2 canaux. Les entrëes dans les Tableaux 4 et 5 qui ont le type de

charge utile "dyn" n'ont pas de type de charge utile statique alloué et ne sont utilisées qu'avec un type de charge utile dynamique. Le type de charge utile 2 avait été alloué à G721 dans la RFC 1890 et à son successeur équivalent G726-32 dans les versions préparatoires de la présente spécification, mais son utilisation est maintenant déconseillée et ce type de charge utile statique est marqué comme réservé à cause du conflit sur l'utilisation des formats de charge utile G726-32 et AAL2-G726-32 (voir au paragraphe 4.5.4). Le type de charge utile 13 indique le format de charge utile de bruit de confort (CN) spécifié dans la RFC 3389 [9]. Le type de charge utile 19 est marqué "réservé" parce que certaines versions préparatoires de la présente spécification allouaient ce numéro à une version antérieure de ce format de charge utile de bruit de confort. La gamme de type de charge utile 72 à 76 est marquée "réservé" afin que les paquets RTCP et RTP puissent être distingués de façon fiable (voir la Section "Résumé des constantes du protocole" de la spécification du protocole RTP).

Les types de charge utile actuellement définis dans ce profil sont alloués à une seule des trois catégories ou types de support : audio seul, vidéo seule, et celles qui combinent audio et vidéo. Les types de supports sont respectivement marqués dans les Tableaux 4 et 5 comme "A", "V" et "AV". Les types de charge utile des différents types de supports NE DEVRONT PAS être entrelacés ou multiplexés au sein d'une même session RTP, mais plusieurs sessions RTP PEUVENT être utilisées en parallèle pour envoyer plusieurs types de supports. Une source RTP PEUT changer les types de charge utile au sein du même type de support durant une session. Voir la section "Multiplexage des sessions RTP" de la RFC 3550 pour des explications supplémentaires.

PT	nom du codage	type de support	débit d'horloge (Hz)	canaux
0	MICU	A	8 000	1
1	réservé	A		
2	réservé	A		
3	GSM	A	8 000	1
4	G723	A	8 000	1
5	DVI4	A	8 000	1
6	DVI4	A	16 000	1
7	LPC	A	8 000	1
8	MICA	A	8 000	1
9	G722	A	8 000	1
10	L16	A	44 100	2
11	L16	A	44 100	1
12	QCELP	A	8 000	1
13	CN	A	8 000	1
14	MPA	A	90 000	(voir le texte)
15	G728	A	8 000	1
16	DVI4	A	11 025	1
17	DVI4	A	22 050	1
18	G729	A	8 000	1
19	réservé	A		
20	non alloué	A		
21	non alloué	A		
22	non alloué	A		
23	non alloué	A		
dyn	G726-40	A	8 000	1
dyn	G726-32	A	8 000	1
dyn	G726-24	A	8 000	1
dyn	G726-16	A	8 000	1
dyn	G729D	A	8 000	1
dyn	G729E	A	8 000	1
dyn	GSM-EFR	A	8 000	1
dyn	L8	A	variable	variable
dyn	RED	A		(voir le texte)
dyn	VDVI	A	variable	1

Tableau 4 : Types de charge utile (PT) pour les codages audio

PT	nom du codage	type de support	débit d'horloge (Hz)
24	non alloué	V	
25	CeLB	V	90 000
26	JPEG	V	90 000
27	non alloué	V	
28	nv	V	90 000

29	non alloué	V	
30	non alloué	V	
31	H261	V	90 000
32	MPV	V	90 000
33	MP2T	AV	90 000
34	H263	V	90 000
35-71	non alloué	?	
72-76	réservé	N/A	N/A
77-95	non alloué	?	
96-127	dynamique	?	
dyn	H263-1998	V	90 000

Tableau 5 : Types de charge utile (PT) pour les codages vidéo et combinés

Les participants à une session se mettent d'accord, par des mécanismes qui sortent du domaine d'application de la présente spécification, sur l'ensemble de types de charge utile permis dans une session donnée. Cet ensemble PEUT, par exemple, être défini par les capacités des applications utilisées, négociées par un protocole de contrôle de conférence ou établies par accord entre les personnes qui participent.

Les applications audio qui fonctionnent sous le présent profil DEVRAIENT, au minimum, être capables d'envoyer et/ou recevoir les types de charge utile 0 (MICU) et 5 (DVI4). Cela permet l'interopérabilité sans négociation de format et assure une négociation réussie avec un protocole de contrôle de conférence.

7. RTP sur TCP et protocoles similaires de flux d'octets

Dans des circonstances particulières, il peut être nécessaire de porter RTP dans des protocoles qui offrent des flux d'octets abstraits, comme TCP, éventuellement multiplexés avec d'autres données. L'application DOIT définir sa propre méthode de délimitation des paquets RTP et RTCP (RTSP [23] fournit un exemple d'une spécification d'encapsulation).

8. Allocation d'accès

Comme spécifié dans la définition du protocole RTP, les données RTP DEVRAIENT être portées sur un numéro d'accès UDP pair et les paquets RTCP correspondants DEVRAIENT être portés sur le numéro d'accès supérieur (impair) suivant.

Les applications qui fonctionnent sous le présent profil PEUVENT utiliser une telle paire d'accès UDP. Par exemple, la paire d'accès PEUT être allouée de façon aléatoire par un programme de gestion de session. Une seule paire fixe de numéros d'accès ne peut pas être exigée parce que plusieurs applications utilisant ce profil vont vraisemblablement fonctionner sur le même hôte, et il y a des systèmes d'exploitation qui ne permettent pas que plusieurs processus utilisent le même accès UDP avec des adresses de diffusion groupée différentes.

Cependant, les numéros d'accès 5004 et 5005 ont été enregistrés pour être utilisés avec ce profil pour les applications qui choisissent de les utiliser comme paire par défaut. Les applications qui fonctionnent sous plusieurs profils PEUVENT utiliser cette paire d'accès comme indication pour choisir ce profil si elle ne sont pas soumises à la contrainte du paragraphe précédent. Les applications n'ont pas besoin d'avoir une paire par défaut et PEUT exiger que la paire d'accès soit explicitement spécifiée. Les numéros d'accès particuliers ont été choisis dans la gamme au dessus de 5000 pour s'accommoder de la pratique de l'allocation des numéros d'accès dans certaines versions du système d'exploitation Unix où les numéros d'accès en dessous de 1024 ne peuvent être utilisés que par des processus privilégiés et les numéros d'accès entre 1024 et 5000 sont automatiquement alloués par le système d'exploitation.

9. Changements depuis la RFC 1890

La présente RFC révisé la RFC 1890. Elle est essentiellement rétro compatible avec la RFC 1890 sauf pour les fonctions retirées parce que on n'a pas trouvé deux mises en œuvre interopérables. Les ajouts à la RFC 1890 codifient les pratiques existantes dans l'utilisation des formats de charge utile sous ce profil. Comme ce profil peut être utilisé sans aucun des formats de charge utile énumérés ici, l'ajout de nouveaux formats de charge utile dans cette révision n'affecte pas la rétro compatibilité. Les modifications sont récapitulées ci-dessous, réparties en changements fonctionnels et non fonctionnels.

Changements fonctionnels :

- o La Section 11, "Considérations relatives à l'IANA" a été ajoutée pour spécifier l'enregistrement du nom de ce profil. Elle fait aussi référence à une nouvelle Section 3 "Enregistrement de codages supplémentaires" qui établit pour politique qu'aucun enregistrement supplémentaires de types de charge utile statique ne sera fait pour ce profil au delà de ceux ajoutés par cette révision et inclus dans les Tableaux 4 et 5. Des noms de codages supplémentaires peuvent à la place être enregistrés comme sous-types MIME pour se lier à des types de charge utile dynamiques. Des références non normatives ont été ajoutées à la RFC 3555 [7] où les sous-types MIME pour tous les formats de charge utile énumérés sont enregistrés, certains avec des paramètres facultatifs pour l'utilisation des formats de charge utile.
- o Les types de charge utile 4, 16, 17 et 34 ont été ajoutés pour incorporer les enregistrements de l'IANA faits depuis la publication de la RFC 1890, ainsi que les descriptions de formats de charge utile correspondants pour G723 et H263.
- o Suite à la discussion du groupe de travail, les types de charge utile statiques 12 et 18 ont été ajoutés ainsi que les descriptions de format de charge utile correspondantes pour QCELP et G729. Le type de charge utile statique 13 a été alloué au format de charge utile Bruit de confort (CN, *Comfort Noise*) défini dans la RFC 3389. Le type de charge utile 19 a été marqué réservé parce qu'il a été temporairement alloué à une précédente version de bruit de confort présente dans des versions antérieures de ce document.
- o Le format de charge utile pour G721 a été renommé G726-32 suivant le changement de numérotation par l'UIT-T, et la description du format de charge utile pour G726 a été étendue pour inclure les débits de données -16, -24 et -40. En raison d'une confusion concernant les versions préparatoires de ce document, certaines mises en œuvre de ce format de charge utile G726 empaquetaient les échantillons en octets en commençant par le bit de poids fort au lieu du bit de moindre poids spécifié ici. Pour résoudre partiellement cette incompatibilité, de nouveaux formats de charge utile nommés AAL2-G726-16, -24, -32 et -40 seront spécifiés dans un document distinct (voir la note du paragraphe 4.5.4) et l'utilisation du type de charge utile statique 2 est déconseillé comme expliqué à la Section 6.
- o Les formats de charge utile G729D et G729E ont été ajoutés suite à l'ajout par l'UIT-T des Annexes D et E à la Recommandation G.729. Des listes ont été ajoutées pour les formats de charge utile GSM-EFR, RED, et H263-1998 publiés dans d'autres documents postérieurs à la RFC 1890. Ces formats de charge utile additionnels ne sont référencés que par des numéros de type de charge utile dynamique.
- o Les descriptions des formats de charge utile pour G722, G728, GSM, VDVI ont été développées.
- o Le format de charge utile pour l'audio 1016 a été retiré et son allocation de type de charge utile statique 1 a été marqué "réservé" parce que deux mises en œuvre interopérables n'ont pas été trouvées.
- o Les exigences pour le contrôle de l'encombrement ont été ajoutées à la Section 2.
- o Le présent profil suit la suggestion de la spécification RTP révisée de spécifier séparément la bande passante RTCP et la bande passante de session ainsi que pour les envoyeurs actifs et les receveurs passifs.
- o La transposition de la chaîne de phrase de passe d'utilisateur en clé de chiffrement a été supprimée de la Section 2 parce que deux mises en œuvre interopérables n'ont pas été trouvées.
- o La convention de rangement des échantillons de "quadriphonie" pour l'audio sur quatre canaux a été retirée pour éliminer une ambiguïté notée au paragraphe 4.1.

Changements non fonctionnels :

- o Au paragraphe 4.1, il est maintenant déclaré explicitement que la suppression de silence est permise pour tous les formats de charge utile audio. (Cela avait toujours été le cas et découle d'un aspect fondamental de la conception de RTP et des motivations du paquet audio, mais n'était pas déclaré explicitement auparavant.) L'utilisation du bruit de confort est aussi expliqué.
- o Au paragraphe 4.1, le niveau d'exigence pour régler le bit marqueur sur le premier paquet après le silence en audio a été changé de "est" en "DEVRAIT être", et précisé que le bit marqueur n'est établi que lorsque les paquets sont intentionnellement non envoyés.
- o De même, du texte a été ajouté pour spécifier que le bit marqueur DEVRAIT être établi à un sur le dernier paquet d'une trame vidéo, et que les trames vidéo sont distinguées par leurs horodatages.
- o Les RFC de référence ont été ajoutées pour les formats de charge utile publiés après la RFC 1890.

- o Les sections de considérations pour la sécurité et de droits de reproduction ont été ajoutées.
- o Conformément à ce que dit Peter Hoddie de Apple, seuls les Macintosh d'avant 1994 utilisaient le débit de 22254,54 et pas le 11127,27 rate, de sorte que ce dernier a été abandonné de la discussion sur les fréquences d'échantillonnage suggérées.
- o Le Tableau 1 a été corrigé pour déplacer certaines valeurs de la colonne "ms/paquet" à la colonne "ms/paquet par défaut" à laquelle elles appartiennent.
- o Comme la "Interactive Multimedia Association" a cessé de fonctionner, un ressource de remplacement a été fournie pour référencer les documents IMA.
- o Une note a été ajoutée sur G722 pour préciser une discordance entre le taux réel d'échantillonnage et le débit d'horloge de l'horodatage RTP.
- o De petites précisions textuelles ont été faites à divers endroits, certaines en réponse à des questions de lecteurs. En particulier :
 - Une définition de "type de support" est donnée en 1.1 pour expliquer plus clairement le multiplexage de sessions RTP à la Section 6 par rapport au multiplexage de plusieurs supports.
 - L'explication de la façon de déterminer dans un paquet le nombre de trames audio à partir de la longueur a été développée.
 - La description de l'allocation de bande passante aux éléments de SDES est donnée.
 - Une note a été ajoutée à la convention spécifiée au paragraphe 4.1 sur l'ordre des canaux qui peut être outrepassé par la spécification d'un codage ou format de charge utile particulier.
 - Les termes DOIT, DEVRAIT, PEUT, etc. sont utilisés comme défini dans la RFC 2119.
- o Un second auteur du document a été ajouté.

10. Considérations pour la sécurité

Les mises en œuvre qui utilisent le profil défini dans cette spécification sont sujettes aux considérations sur la sécurité exposées dans la spécification RTP [1]. Le présent profil ne spécifie aucun service de sécurité différent. La principale fonction de ce profil est de faire la liste d'un ensemble de codages de compression de données pour les supports audio et vidéo.

La confidentialité des flux de supports est réalisée par le chiffrement. Comme la compression des données utilisée avec les formats de charge utile décrits dans ce profil est appliquée de bout en bout, le chiffrement peut être effectué après la compression de sorte qu'il n'y a pas de conflit entre les deux opérations.

Une menace potentielle de déni de service existe pour les codages de données qui utilisent les techniques de compression qui ont une charge de calcul non uniforme à l'extrémité de réception. L'attaquant peut injecter dans le flux des datagrammes pathogènes complexes à décoder qui causent la surcharge du receveur.

Comme avec tout protocole fondé sur IP, dans certaines circonstances un receveur peut être surchargé simplement par la réception de paquets trop nombreux, désirés ou indésirables. L'authentification de couche réseau PEUT être utilisée pour éliminer des paquets provenant de sources non désirées, mais le coût de traitement de l'authentification elle-même peut être trop élevé. Dans un environnement de diffusion groupée, l'élagage de source est mis en œuvre dans IGMPv3 (RFC 3376) [24] et dans les protocoles d'acheminement de diffusion groupée pour permettre à un receveur de choisir quelles sources sont autorisées à le joindre.

11. Considérations relatives à l'IANA

La spécification RTP établit un registre des noms de profils à utiliser par les protocoles de contrôle de niveau supérieur, tels que le protocole de description de session (SDP), RFC 2327 [6], pour se référer aux méthodes de transport. Ce profil enregistre le nom "RTP/AVP".

La Section 3 établit la politique selon laquelle aucun enregistrement supplémentaire de type de charge utile RTP statique ne sera fait pour ce profil en plus de ceux ajoutés dans la révision de ce document et inclus dans les tableaux 4 et 5. L'IANA peut faire référence à cette section pour refuser d'accepter toute demande d'enregistrement supplémentaire. Dans les

tableaux 4 et 5, noter que les types 1 et 2 ont été marqués comme réservés et que l'ensemble de types de charge utile "dyn" inclus a été mis à jour. Ces changements sont expliqués aux Sections 6 et 9.

12. Références

(Les liens hypertextes dans le titre sont sur la traduction française éventuelle.)

12.1 Références normatives

- [1] H. Schulzrinne, S. Casner, R. Frederick et V. Jacobson, "RTP : un protocole de [transport pour les applications](#) en temps réel", STD 64, [RFC 3550](#), juillet 2003.
- [2] S. Bradner, "[Mots clés à utiliser](#) dans les RFC pour indiquer les niveaux d'exigence", BCP 14, [RFC 2119](#), mars 1997.
- [3] Apple Computer, "Audio Interchange File Format AIFF-C", août 1991. (aussi à <ftp://ftp.sgi.com/sgi/aiff-c.9.26.91.ps.Z>).

12.2 Références pour information

- [4] R. Braden, D. Clark et S. Shenker, "Intégration de services dans l'architecture de l'Internet : Généralités", RFC 1633, juin 1994. *(Information)*
- [5] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang et W. Weiss, "Architecture pour services différenciés", RFC 2475, décembre 1998. *(Information, mise à jour par RFC 3260)*
- [6] M. Handley et V. Jacobson, "SDP : Protocole de description de session", RFC 2327, avril 1998. *(Rendue obsolète par la RFC 4566)*
- [7] S. Casner et P. Hoschka, "Enregistrement de type MIME pour les types de charge utile RTP", RFC 3555, juillet 2003. *(Rendue obsolète par les RFC 4855/4856)*
- [8] N. Freed, J. Klensin et J. Postel, "Extensions multi-objets de la messagerie Internet (MIME) Partie 4 : Procédures d'enregistrement", BCP 13, RFC 2048, novembre 1996. *(Rendue obsolète par RFC 4288/4289)*
- [9] R. Zopf, "Charge utile du protocole de transport en temps réel (RTP) pour le bruit de confort", [RFC 3389](#), septembre 2002. *(Proposition de norme)*
- [10] Deleam, D. and J.-P. Petit, "Real-time implementations of the recent ITU-T low bit rate speech coders on the TI TMS320C54X DSP: results, methodology, and applications", dans Proc. of International Conference on Signal Processing, Technology, and Applications (ICSPAT), (Boston, Massachusetts), pp. 1656--1660, octobre 1996.
- [11] M. Mouly et M.-B. Pautet, "Le système GSM pour les communications entre les mobiles", Lassay-les-Chateaux, France, Europe Media Duplication, 1993.
- [12] Degener, J., "Digital Speech Compression", Dr. Dobb's Journal, décembre 1994.
- [13] Redl, S., Weber, M. and M. Oliphant, "An Introduction to GSM", Boston: Artech House, 1995.
- [14] D. Hoffman, G. Fernando, V. Goyal et M. Civanlar, "Format de charge utile RTP pour la vidéo MPEG1/MPEG2", RFC 2250, janvier 1998. *(Proposition de norme)*
- [15] Jayant, N. and P. Noll, "Digital Coding of Waveforms--Principles and Applications to Speech and Video", Englewood Cliffs, New Jersey: Prentice-Hall, 1984.
- [16] K. McKay, "Format de charge utile RTP pour les flux audio PureVoice(tm)", RFC 2658, août 1999. *(Prop. de norme)*
- [17] C. Perkins, I. Kouvelas, O. Hodson, V. Hardman, M. Handley, J. Bolot, A. Vega-Garcia et S. Fosse-Parisis, "Charge utile RTP pour données audio redondantes", RFC 2198, septembre 1997. *(Proposition de norme)*
- [18] M. Speer et D. Hoffman, "Format de charge utile RTP pour les codages vidéo CellB de Sun", RFC 2029, octobre 1996. *(Proposition de norme)*

- [19] L. Berc, W. Fenner, R. Frederick, S. McCanne et P. Stewart, "Format de charge utile RTP pour la vidéo compressée en JPEG", RFC 2435, octobre 1998. (*Proposition de norme*)
- [20] T. Turlletti et C. Huitema, "Format de charge utile RTP pour les flux vidéo H.261", RFC 2032, octobre 1996. (*Rendue obsolète par la RFC 4587*)
- [21] C. Zhu, "Format de charge utile RTP pour les flux vidéo H.263", RFC 2190, septembre 1997. (*Historique*)
- [22] C. Bormann, L. Cline, G. Deisher, T. Gardos, C. Maciocco, D. Newell, J. Ott, G. Sullivan, S. Wenger et C. Zhu, "Format de charge utile RTP pour la version 1998 de la Rec. UIT-T H.263 Vidéo (H.263+)", RFC 2429, octobre 1998. (*Rendue obsolète par la RFC 4629*)
- [23] H. Schulzrinne, A. Rao et R. Lanphier, "Protocole de flux directs en temps réel (RTSP)", [RFC 2326](#), avril 1998. (*Proposition de norme*)
- [24] B. Cain, S. Deering, I. Kouvelas, B. Fenner et A. Thyagarajan, "Protocole Internet de gestion de groupe, version 3", [RFC 3376](#), octobre 2002. (*Mise à jour par la RFC 4004*)

13. Localisation actuelle des ressources concernées

Note : Plusieurs sections ci dessous se réfèrent à la Bibliothèque d'outils logiciels de l'UIT-T (STL, *Software Tool Library*). Celle-ci est disponible auprès du Service des ventes de l'UIT, Place des Nations, CH-1211 Genève 20, Suisse (voir aussi <http://www.itu.int>). Le STL de l'UIT-T est couvert par un brevet défini dans la Recommandation UIT-T G.191, "Outils logiciels pour la normalisation des codages de parole et audio".

DVI4

Une copie d'archive du document "IMA Recommended Practices for Enhancing Digital Audio Compatibility in Multimedia Systems (version 3.0)", qui décrit l'algorithme ADPCM de l'IMA, est disponible à :

<http://www.cs.columbia.edu/~hgs/audio/dvi/>

Une mise en œuvre est disponible auprès de Jack Jansen à <ftp://ftp.cwi.nl/local/pub/audio/adpcm.shar>

G722

Une mise en œuvre de l'algorithme de G.722 est disponible au titre de la STL de l'UIT-T, décrite ci-dessus.

G723

La mise en œuvre de référence du code C qui définit l'algorithme de G.723.1 et ses Annexes A, B, et C sont disponibles dans la Recommandation G.723.1 auprès du service des ventes de l'UIT, à l'adresse ci-dessus. L'algorithme et le code C sont tous deux couverts par un brevet spécifique. Le Secrétariat de l'UIT doit être contacté pour obtenir les informations.

G726

G726 est spécifié dans la Recommandation UIT-T G.726, "Modulation par impulsions codées différentielles adaptatives à 40, 32, 24, et 16 kbit/s (ADPCM)". Une mise en œuvre de l'algorithme de G.726 est disponible au titre de la STL de l'UIT-T, décrite plus haut.

G729

La mise en œuvre de référence du code C qui définit l'algorithme de G.729 et ses Annexes A à I sont disponibles au titre de la Recommandation G.729 auprès du service des ventes de l'UIT, cité ci-dessus. L'Annexe I contient le code source C intégré pour tous les modes de fonctionnement de G.729. L'algorithme G.729 et le code C associé sont couverts par un brevet spécifique. Les informations de contact pour obtenir la licence sont disponibles auprès du Secrétariat de l'UIT.

GSM

Une mise en œuvre de référence a été rédigée par Carsten Bormann et Jutta Degener (alors à TU Berlin, Allemagne). Elle est disponible à <http://www.dmn.tzi.org/software/gsm/>

Bien que l'algorithme RPE-LTP ne soit pas une norme de l'UIT-T, il y a une mise en œuvre de code C de l'algorithme RPE-LTP disponible au titre de la STL de l'UIT-T. C'est une adaptation de la version de l'Université Télécom de Berlin.

LPC

Une mise en œuvre est disponible à <ftp://parcftp.xerox.com/pub/net-research/lpc.tar.Z>

MICU, MICA

Une mise en œuvre de ces algorithmes est disponible au titre de la STL de IUT-T, décrite plus haut.

14. Remerciements

Les commentaires et la relecture attentive de Simao Campos, Richard Cox et des participants au groupe de travail AVT ont droit à notre pleine gratitude. La description du GSM a été adaptée de l'Accord pour la mise en œuvre de l'interopérabilité des services du Forum IMTC pour la voix sur IP (janvier 1997). Fred Burg et Terry Lyons ont aidé pour la description de G.729.

15. Déclaration de droits de propriété intellectuelle

L'IETF ne prend pas position sur la validité et la portée de tout droit de propriété intellectuelle ou autres droits qui pourraient être revendiqués au titre de la mise en œuvre ou l'utilisation de la technologie décrite dans le présent document ou sur la mesure dans laquelle toute licence sur de tels droits pourrait être ou n'être pas disponible ; pas plus qu'elle ne prétend avoir accompli aucun effort pour identifier de tels droits. Les informations sur les procédures de l'IETF au sujet des droits dans les documents en cours de normalisation et se rapportant aux normes figurent dans le BCP 11.

Des copies des dépôts d'IPR faits au secrétariat de l'IETF et toutes assurances de disponibilité de licences, ou le résultat de tentatives faites pour obtenir une licence ou permission générale d'utilisation de tels droits de propriété par ceux qui mettent en œuvre ou utilisent la présente spécification peuvent être obtenues sur le répertoire en ligne des IPR de l'IETF à <http://www.ietf.org/ipr>.

L'IETF invite toute partie intéressée à porter son attention sur tous copyrights, brevet ou applications de brevets, ou autres droits de propriété qui pourraient recouvrir la technologie qui pourrait être nécessaire pour mettre en œuvre la présente norme. Prière d'adresser les informations au directeur exécutif de l'IETF.

16. Adresse des auteurs

Henning Schulzrinne
Department of Computer Science
Columbia University
1214 Amsterdam Avenue
New York, NY 10027
United States
mél : schulzrinne@cs.columbia.edu

Stephen L. Casner
Packet Design
3400 Hillview Avenue, Building 3
Palo Alto, CA 94304
United States
mél : casner@acm.org

17. Déclaration complète des droits de reproduction

Copyright (C) The Internet Society (2003). Tous droits réservés

Le présent document et ses traductions peuvent être copiés et fournis aux tiers, et les travaux dérivés qui les commentent ou les expliquent ou aident à leur mise en œuvre peuvent être préparés, copiés, publiés et distribués, en tout ou partie, sans restriction d'aucune sorte, pourvu que la déclaration de copyright ci-dessus et le présent paragraphe soient inclus dans toutes ces copies et travaux dérivés. Cependant, le présent document lui-même ne peut être modifié d'aucune façon, en particulier en retirant la notice de copyright ou les références à la Internet Society ou aux autres organisations Internet, excepté autant qu'il est nécessaire pour les besoins du développement des normes Internet, auquel cas les procédures de copyright définies dans les procédures des normes Internet doivent être suivies, ou pour les besoins de la traduction dans d'autres langues que l'anglais.

Les permissions limitées accordées ci-dessus sont perpétuelles et ne seront pas révoquées par la Internet Society ou successeurs ou ayant droits.

Le présent document et les informations y contenues sont fournies sur une base "EN L'ÉTAT" et le contributeur, l'organisation qu'il ou elle représente ou qui le/la finance (s'il en est), la INTERNET SOCIETY et la INTERNET ENGINEERING TASK FORCE déclinent toutes garanties, exprimées ou implicites, y compris mais non limitées à toute garantie que l'utilisation des informations ci-encloses ne violent aucun droit ou aucune garantie implicite de commercialisation ou d'aptitude à un objet particulier.

Remerciement

Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.