

RTP et RTCP

EFORT

<http://www.efort.com>

Pour transporter la voix ou la vidéo sur IP, le protocole IP (Internet Protocol) au niveau 3 et le protocole UDP (User Datagram Protocol) au niveau 4 sont utilisés. En effet, si TCP (Transmission Control Protocol) présente l'avantage de gérer un transfert fiable (renvoi des paquets IP en cas d'erreur), il est malheureusement incompatible avec un flux temps-réel dans la mesure où les mécanismes de TCP prévoient une réduction automatique du débit accordé à l'émetteur en cas de congestion du réseau et une remontée lente vers le débit nominal (ceci afin de protéger le réseau de soubresauts d'émetteurs qui chercheraient tous en même temps à tirer parti de la bande passante disponible).

Mais ces deux protocoles UDP et IP ne suffisent pas à assurer le transport de la voix. De fait, UDP est un protocole sans correction d'erreur, et à aucun moment l'arrivée des paquets dans leur ordre d'émission est assurée. Pour le transport de données temps réel telles que la voix ou la vidéo, il est nécessaire d'utiliser deux protocoles supplémentaires : RTP (Real-Time transport Protocol) et RTCP (RTP Control Protocol).

RTP et RTCP sont des protocoles qui se situent au niveau de l'application et s'appuient sur le protocole de transport UDP. RTP et RTCP peuvent utiliser aussi bien le mode Unicast (point à point) que le mode Multicast (multipoint).

RTP et RTCP utilisent des ports différents. RTP utilise un numéro de port pair, et RTCP le numéro de port impair qui suit directement. Lorsqu'une session RTP est ouverte, alors une session RTCP est aussi ouverte de manière implicite.

Les numéros de port utilisés par RTP et RTCP sont compris entre 1025 et 65535. Les ports RTP et RTCP par défaut sont respectivement 5004 et 5005.

Le paragraphe 1 introduit le protocole RTP. Le paragraphe 2 présente les entités Mixer et Translator pouvant être utilisées lors d'échanges de paquets RTP. Le paragraphe 3 décrit la structure des paquets RTP. Enfin le paragraphe 4 détaille le protocole RTCP, son fonctionnement et les paquets RTCP.

1. RTP (Real-Time Transport Protocol)

Le but de RTP est de fournir un moyen uniforme de transmettre sur IP des données soumises à des contraintes de temps réel (audio, vidéo, etc.).

RTP permet :

- d'identifier le type de l'information transportée,
- d'ajouter des marqueurs temporels permettant d'indiquer l'instant d'émission du paquet. L'application destinataire peut alors synchroniser les flux et mesurer les délais et la gigue.
- d'inclure des numéros de séquence à l'information transportée afin de détecter l'occurrence de paquets perdus et de délivrer les paquets en séquence à l'application destinataire.

De plus, RTP peut être véhiculé par des paquets multicast afin d'acheminer des conversations vers des destinataires multiples.

Mais, RTP n'a pas été conçu pour effectuer des réservations de ressources ou contrôler la qualité de service et ne garantit pas la livraison du paquet à l'arrivée.

2. Mixer et Translator

Le mixer (*mixeur*) est une application qui reçoit des flux de données de plusieurs sources qu'on appelle SSRCs (Synchronization Sources), en modifie éventuellement le format et

renvoie un seul flux de données agrégé. Là encore, dans une application de conférence, il est probablement économique de regrouper dès que possible les flux sonores provenant de plusieurs sources, puisque, de toutes façons, ils seront mélangés chez les destinataires. En faisant ce travail, le mixer se comporte comme une source particulière (SSRC) qui regroupe les données de plusieurs autres sources qui étaient des SSRCs et qui deviennent des CSRC (*Contributing Source*). Les paquets émis par une quelconque source (terminal, mixer) contiennent une information d'identification de la SSRC qui les émet. Ces paquets peuvent également contenir l'identification des CSRC qui sont les sources réelles des informations lorsqu'il s'agit d'un paquet construit par transformation.

Dans l'exemple représenté à la figure 1, un mixer reçoit des paquets de trois sources (SSRC). Il peut réaliser des conversions de format, mixe le contenu en fonction de l'application et produit un nouveau paquet RTP qu'il relaye à la destination.

Le mixer fournit la synchronisation pour le nouveau flux et s'identifie comme la nouvelle source SSRC.

Les trois sources origines sont déclarées comme CSRC dans le paquet RTP composé.

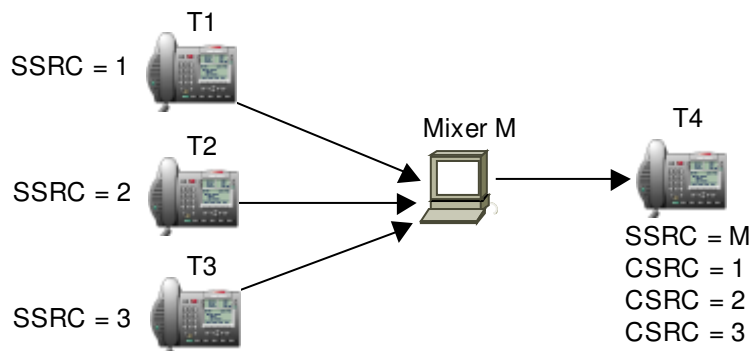


Figure 1 : Mixer RTP

Le traducteur (traducteur) est une application qui transmet les paquets RTP qu'elle reçoit sans changer l'identificateur de SSRC (à l'inverse de ce que fait le mixer). Un traducteur permet par exemple de changer le codage d'une donnée ou le débit, ou encore de traiter les problèmes de sécurité (*firewalls*) à la frontière d'un réseau privé.

Si un terminal utilisant un codage PCM μ -law à 64 kbit/s souhaite établir une session avec un terminal supportant un codage G.726 ADPCM à 32 kbit/s, alors il est nécessaire d'interfacer les deux terminaux à travers un traducteur.

Les fonctions Mixer et Translator sont généralement mises en œuvre dans un Gateway.

3. Les données RTP

RTP transporte les signaux audio ou vidéo encodés à l'aide de paquets RTP contenant un header RTP (en-tête) suivi de ces signaux audio ou vidéo.

Un paquet RTP est soumis à la couche UDP qui y rajoute un en-tête UDP. L'ensemble est soumis à la couche IP qui y agrège un en-tête IP. Le datagramme IP est alors routé à la destination. A la réception, le paquet est délivré à la bonne application.

3.1. Codage des signaux

Il est essentiel compte tenu du nombre important de normes de codage de signaux audio ou vidéo d'inclure un mécanisme à RTP afin de permettre au destinataire de connaître le codage utilisé et ainsi pouvoir décoder correctement. RTP réalise cette fonction grâce au numéro de type de contenu (payload type number) dans le header (en-tête) RTP. Les

numéros de type de contenu sont spécifiés dans le RFC 3551 (RTP Profile for Audio and Video Conferences with Minimal Control) et sont listés dans le tableau 1.

Type Payload	Codec	Fréquence (Hz)	Description
0	PCMU	8000	ITU G.711 PCM μ -Law audio 64 kbit/s
1	1016	8000	CELP Audi 4.8 kbit/s
2	G721	8000	ITU G.721 ADPCM Audio 32 kbit/s
3	GSM	8000	European GSM Audio 13 kbit/s
5	DVI4	8000	DVI ADPCM Audio 32kbit/s
6	DVI4	16000	DVI ADPCM Audio 64kbit/s
7	LPC	8000	Experimental LPC Audio
8	PCMA	8000	ITU G.711 PCM A-Law audio 64 kbit/s
9	G722	8000	ITU G.722 Audio
10	L16	44100	Linear 16 bit Audio 705,6 kbit/s
11	L16	44100	Linear 16 bit Stereo Audio 1411.2 kbit/s
14	MPA	90000	MPEG-I ou MPEG-II Audio
15	G728	8000	ITU G.728 Audio 16kbit/s
25	CELB	90000	CelB Video
26	JBEG	90000	JBEG Video
28	NV	90000	Nv Video
31	H261	90000	ITU H.261 Video
32	MPV	90000	MPEG-I et MPEG-II Video
33	MP2T	90000	MPEG-II transport stream Video

Tableau 1 : Les types de payload pour l'encodage de signaux audio et vidéo

3.2. Encodage PCM

Le code du son avec une qualité de type téléphonique est basé sur le PCM μ -law ou parfois a-law. La norme μ -law était celle en vigueur sur le réseau téléphonique américain, la norme a-law celle en vigueur sur le réseau téléphonique européen. Dans cette norme, le son est codé sur 14 bits avec une fréquence de 8000 Hertz, puis chaque échantillon est ramené à une largeur de 8 bits à l'aide d'une table de translation. C'est la différence entre ces tables de translation qui sépare les deux codages μ -law et a-law. Cette fréquence d'échantillonnage fournit une fréquence audible couvrant la place de 300 Hertz à 3500 Hertz. Avec des échantillons de 8 bits, le flux généré atteint un débit de 64 kbit/s.

3.3. Encodage ADPCM

Une compression ADPCM par codage différentiel a été proposée afin de réduire le débit généré. Celui-ci passe à 32 kbit/s. Cette compression altère très faiblement la qualité sonore, tout au plus, limite-t-elle la dynamique dans les aigus.

3.4. Format du paquet RTP

L'en-tête d'un paquet RTP (RTP header) est obligatoirement constitué de 12 octets, éventuellement suivi d'une liste d'identificateurs de sources contributives CSRCs dans le cas d'un mixer. Cet en-tête précède le "payload" qui représente les données utiles (Figure 2).

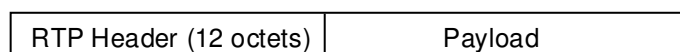


Figure 2 : Format du paquet RTP

- Version (V) (2 bits) : Ce champ indique le numéro de version RTP utilisée. La version actuelle est la version 2.

- Padding (P) (1 bit) : Si positionné à 1, ce champ signifie que le champ des données (payload) a une partie de bourrage. Rappelons que la longueur des données doit être un multiple de 4 octets, d'où la nécessité d'octets de bourrage, notamment pour le dernier paquet. Le dernier octet de cette partie de bourrage indique le nombre d'octets de bourrage à ignorer.
- Extension (X) (1 bit) : Si positionné à 1, ce champ indique que l'en-tête fixe a une partie d'en-tête supplémentaire.
- CSRC Count (CC) (4 bits) : Ce champ contient le nombre d'identifiants CSRC qui suivent l'en-tête fixe, c'est à dire le nombre de sources contributives liées à ce paquet.
- Marker (M) (1 bit) : Il s'agit d'un bit de signalisation. Sa signification dépend des données transportées.
- Payload type (PT) (7 bits) : Ce champ identifie le type de contenu (audio, vidéo, etc.) qui représente le type de codage d'information véhiculé dans le paquet. La liste est présentée au tableau 8.1.
- Sequence Number (16 bits) : La valeur de ce champ est incrémentée de 1 à chaque paquet RTP envoyé alors que sa valeur initiale est aléatoire. Ce champ permet de détecter des paquets RTP perdus.
- TimeStamp (32 bits) : Un protocole comme RTP utilise des estampilles temporelles pour dater les paquets émis. Ces informations sont la base de calculs permettant d'évaluer le délai et la gigue introduits par un système de communication. La pertinence de ces informations dépend toute entière de la précision et de la synchronisation des horloges utilisées dans les machines qui vont appuyer leurs calculs sur ces valeurs.
- Synchronization Source (SSRC) (32 bits) : Ce champ identifie la source ayant produit le paquet. Au début d'une session, chaque participant choisit un numéro de SSRC. On parle ici de synchronisation car l'échelle de temps installée par la source dans ses paquets va servir de repère aux récepteurs pour restituer l'information correctement.
- Contributing source (CSRC) : 0 à 15 instances de ce champ peuvent être présentes dans l'en-tête du paquet RTP. Le nombre est indiqué par le champ CC. Lorsqu'un flux RTP est le résultat d'une agrégation par un mixeur RTP de plusieurs flux RTP, la liste des SSRCs ayant apporté leur contribution est rajoutée dans l'en-tête des paquets RTP du flux résultant (Figure 3).

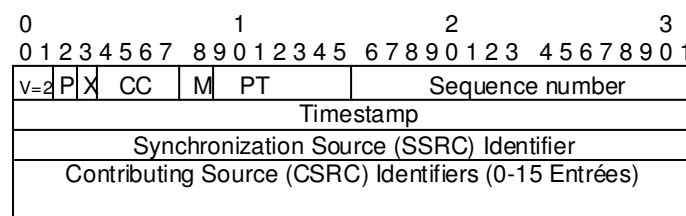


Figure 3 : En-tête RTP

4. RTCP (RTP Control Protocol)

Le protocole RTCP est basé sur des transmissions périodiques de paquets de contrôle par tous les participants dans la session.

C'est un protocole de contrôle des flux RTP, permettant de véhiculer des informations basiques sur les participants d'une session, et sur la qualité de service.

Il existe cinq types différents de paquets RTCP pour chaque type d'information :

- SR (Sender Report) contient des statistiques de transmission et de réception pour les participants qui sont des émetteurs actifs.
- RR (Receiver Report) contient des statistiques de réception pour les participants qui ne sont pas des émetteurs actifs mais récepteurs d'une session.
- SDES (Source Description) décrit la source : nom, email, tél, etc.
- BYE permet à une station d'indiquer la fin de sa participation à une session.
- APP est un paquet de signalisation spécifique à une application.

Le contrôle de flux RTP est réalisé en gardant une évaluation du nombre de participants à une session (sources et récepteurs). A partir de cette évaluation on calcule un intervalle de temps qui sert de période de récurrence à la diffusion des informations SR ou RR suivant le cas. Globalement, les algorithmes de contrôle limitent le volume des informations de contrôle transmises (les données RTCP donc) à 5% du volume global des échanges de la session. Dans ce volume, 25% sont réservées aux informations des sources (messages SR). On garantit ainsi une possibilité de gérer des groupes de grande taille du point de vue du volume d'information échangé. Plus le nombre de participants est élevé, moins précise est la vision qu'a chaque participant de l'état du réseau.

Si l'on considère une session audio avec deux participants, des paquets RTCP peuvent être émis toutes les 5 secondes alors que pour quatre participants, ils sont émis toutes les 10 secondes. Les paquets qui sont le plus fréquemment transmis sont SR et RR.

4.1. Paquet RTCP Sender Report (SR)

Les participants à une session qui à la fois émettent et reçoivent des paquets RTP utilisent les paquets RTCP SR. Le format de ce paquet est présenté à la figure 4.

Le paquet SR contient un header (en-tête), des informations sur l'émetteur, un certain nombre de blocs de rapports de réception et optionnellement une extension spécifique au profil.

La structure est similaire à celle d'un paquet RTP.

L'en-tête contient les champs suivants :

- Version (V) (2 bits) : Ce champ indique le numéro de version RTCP. La valeur de la version courante du protocole RTCP est 2 (10).
- Padding (P) (1 bit) : Positionné à la valeur 1, ce champ indique qu'il y a un bourrage dont la taille est indiquée dans le dernier octet.
- Reception report count (RC) (5 bits) : Ce champ précise le nombre de rapports de réception contenus dans le paquet SR, en considérant un rapport pour chaque source. Un maximum de 31 rapports peut donc être inclus dans le paquet SR.
- Packet type (PT) (8 bits) : Ce champ indique le type de paquet ; il s'agit d'un paquet SR, représenté par la valeur 200.
- Length (16 bits) : Ce champ indique la longueur totale du paquet en mots de 32 bits (en-tête et bourrage compris).

Les informations sur l'émetteur consistent en les champs suivants :

- SSRC of sender (32 bits) : Ce champ précise l'identification de la source spécifique à l'émetteur.
- NTP timestamp (64 bits) : La représentation du temps utilisée par NTP (Network Time Protocol) est assez simple : une date est codée sur 64 bits et mesurée en secondes depuis 0h le premier janvier 1900. La partie entière de cette date exprimée en secondes est codée sur les 32 bits de poids fort (most significant word), la partie fractionnaire (fraction de seconde) sur les 32 bits de poids faible (least significant word). Cette représentation garantit une précision d'à peu près 200 pico-secondes, ce qui est probablement suffisant pour la plupart des applications. Le problème du bouclage à zéro de cette représentation interviendra en 2036. Il sera donc nécessaire de définir une nouvelle version du protocole avant cette date.

RTP timestamp (32 bits) : Ce champ indique le même temps que celui présent le champ NTP Timestamp précédent, mais en utilisant les mêmes unités que celles utilisées pour spécifier la valeur timestamp dans les paquets RTP.

- Sender's packet count (32 bits) : Ce champ indique le nombre total de paquets RTP transmis par l'émetteur depuis le début de la session. Il est réinitialisé dans le contexte d'une session si l'émetteur change d'identificateur SSRC.

- Sender's octet count (32 bits) : Ce champ indique le nombre total d'octets RTP (ne sont considérés que les octets de données utilisateur et non les octets d'en-tête ou de bourrage) transmis par l'émetteur depuis le début de la session. Il est aussi réinitialisé si l'émetteur change d'identificateur de SSRC.

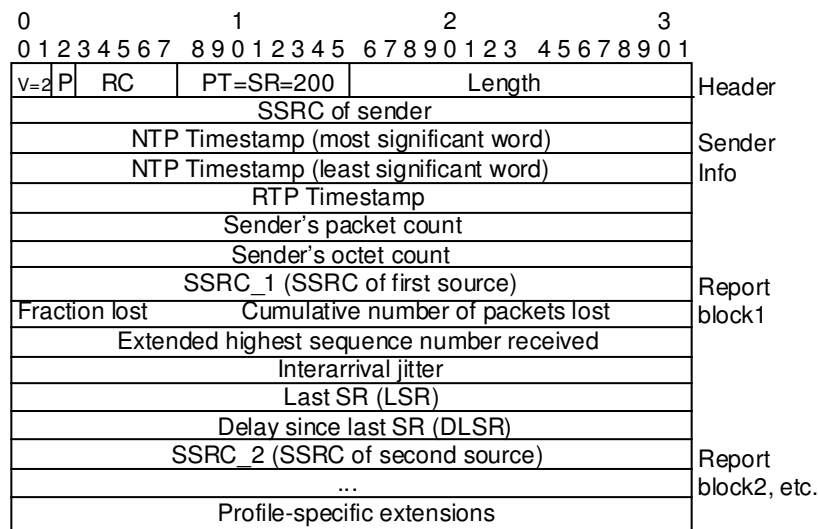


Figure 4 : Paquet RTCP Sender Report

Un ou plusieurs blocs de rapports de réception (RR, Receiver Report) suivent l'information de l'émetteur. Ils fournissent aux autres participants de la session l'information concernant le nombre de paquets RTP qu'ils ont émis et qui ont été reçus avec succès par l'émetteur du paquet SR.

Les champs suivants sont inclus dans chaque bloc RR (Figure 5):

- SSRC_n (32 bits): Ce champ précise l'identification de la source dans la session, qui est concernée par les données incluses dans le bloc RR.
- Fraction lost (8 bits): Ce champ indique la fraction de paquets RTP perdus depuis le dernier rapport émis par ce participant. La fraction représente le rapport entre le nombre de paquets perdus et le nombre de paquets attendus. Le nombre de paquets perdus peut être déduit à partir de l'analyse du numéro de séquence (Sequence Number) de chaque paquet RTP reçu.
- Cumulative number of packets lost (24 bits): Ce champ indique le nombre total de paquets RTP de la source en question qui ont été perdus depuis le début de la session RTP.
- Extended highest sequence number received (32 bits): Ce champ précise le numéro de séquence du dernier paquet RTP reçu depuis cette source SSRC_n.
- Interarrival jitter (32 bits) : Ce champ renseigne sur la variation du délai de transmission des paquets RTP.
- Last SR Timestamp (LSR) (32 bits) : Ce champ représente les 32 bits du milieu du champ NTP Timestamp utilisé dans le dernier paquet SR reçu depuis la source en question. Il considère donc les 16 bits de poids faible de la partie entière de cette date (secondes) et les 16 bits de poids fort de la partie fractionnaire (fractions de seconde). Si aucun paquet RTCP SR n'a encore été reçu, alors la valeur de ce champ est égale à 0.
- Delay Since Last SR (DLSR) (32 bits) : Ce champ représente le délai exprimé en unités de 1/65536 secondes entre l'instant de réception du dernier paquet SR de la source SSRC_n et l'instant d'émission de ce bloc RR. Si aucun paquet SR n'a encore été reçu de la source SSRC_n, la valeur du champ DLSR est positionnée à 0.

4.2. Paquet RTCP Receiver Report (RR)

Le paquet RCTP RR (Receiver Report) est émis par un participant à une session qui reçoit des paquets RTP mais n'en émet pas. Le format du paquet est présenté à la figure 5. Il a une structure similaire au paquet RTCP SR, mais indique la valeur 201 pour le champ payload type et n'inclut pas d'information spécifique à l'émetteur.

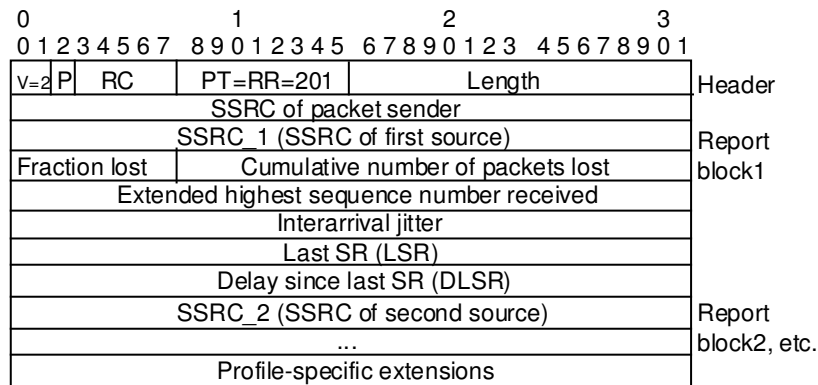


Figure 5 : Paquet RTCP Receiver Report

4.3. Paquet Source Description (SDES)

Le paquet RTCP SDES (Source Description) permet d'identifier les participants à une session et de fournir des informations sur ces participants. La structure du paquet SDES est composée d'un header (en-tête) et de zéro, un ou plusieurs chunks, chacun décrivant la source identifiée dans ce chunk (Figure 6).

L'en-tête comprend :

- Les champs Version Padding et Length qui ont la même signification que ceux du paquet RTCP SR.
- Le champ Packet Type (PT) sur 8 bits qui indique un paquet SDES, représenté par la valeur 202.
- Le champ Source Count (SC) sur 5 bits qui indique le nombre de chunks SSRC/CSRC contenus dans le paquet SDES.

Chaque Chunk contient une valeur de SSRC ou de CSRC suivie par un ou plusieurs identificateurs et par des informations décrivant ce SSRC ou CSRC. Ces informations sont dénommées « SDES items » et peuvent inclure des données telles que le nom, l'adresse Email, le numéro de téléphone, la localisation de l'utilisateur, etc.

Les terminaux doivent envoyer un paquet SDES au début de la session afin que chaque participant soit explicitement identifié. Le mixer combine les paquets SDES de différents participants et produit un paquet composé d'autant de Chunks que de paquets SDES origine.

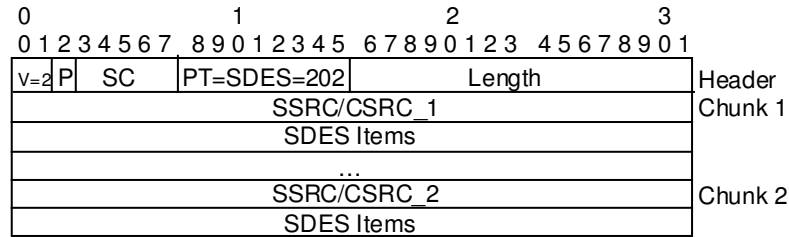


Figure 6 : Paquet RTCP Source Description

4.4. Paquet RTCP BYE

Le paquet RTCP BYE permet d'indiquer qu'une source (un participant à la session) n'est plus active.

Les champs Version, Padding et Length ont la même signification que ceux du paquet RTCP SR.

Le champ Packet Type (PT) indique le type de paquet RTCP. il s'agit d'un paquet BYE, représenté par la valeur 203.

Le champ Source Count (SC) indique le nombre d'identificateurs SSRC/CSRC inclus dans ce paquet.

Si un paquet BYE est reçu par un mixer, il le relaye avec les identificateurs SSRC/CSRC inchangés. Si par contre le mixer doit arrêter de fonctionner, il doit alors émettre un paquet BYE listant toutes les sources contributives qu'il prend en charge, ainsi que son propre identificateur SSRC.

Enfin, le champ optionnel Reason for leaving indique, sous la forme d'une chaîne de caractères, la raison de l'envoi du paquet BYE. Ce champ est précédé par sa longueur.

A la différence du paquet SDES, il n'est pas possible de mixer plusieurs paquets BYE en un seul paquet BYE composé.

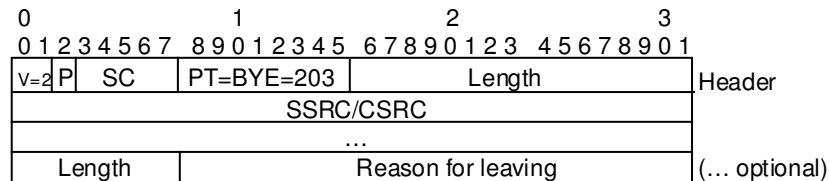


Figure 7 : Paquet RTCP BYE

4.5. Paquet RTCP APP

Le paquet RTCP APP est un paquet de signalisation spécifique aux applications (Figure 8).

Les champs Version, Padding et Length ont la même signification que ceux du paquet RTCP SR.

Le champ subtype (5 bits) peut être utilisé afin de définir un sous-type identifiant un ensemble de paquets APP.

Le champ name affecte un nom unique (4 caractères) pour le sous-type.

Les données dépendant de l'application sont incluses dans le champ « Application-dependent data ».

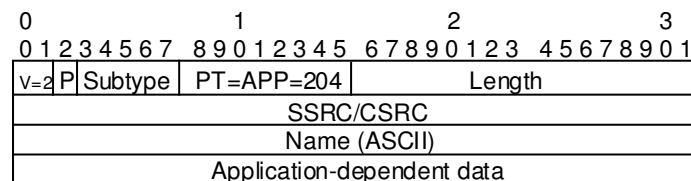


Figure 8 : Paquet RTCP APP

5. Conclusion

Les protocoles RTP et RTCP sont adaptés pour la transmission de données temps réel. Cependant, ils fonctionnent en stratégie bout à bout et donc ne peuvent contrôler l'élément principal de la communication : le réseau. Pourtant, quelques soient les efforts d'adaptation des émetteurs, ou les moyens mis en œuvre par les récepteurs, c'est au cœur du réseau que les dysfonctionnements critiques sont générés. Le protocole Internet a été volontairement pensé pour reporter l'intelligence vers les systèmes d'extrémité. C'est cette simplicité qui a conduit au succès d'Internet. Le protocole RSVP (Resource Reservation Protocol) défini par l'IETF a été développé afin de remédier à ces dysfonctionnements et ainsi améliorer les transmissions temps réel.

Les protocoles RTP et RTCP sont principalement utilisés en visioconférence, où les participants sont tour à tour, émetteurs ou récepteurs. Pour le transport de la voix, ils permettent une transmission correcte sur des réseaux bien ciblés. C'est-à-dire, des réseaux qui implémentent une qualité de service adaptée (ATM). Il est aussi possible de s'appuyer sur des réseaux bien dimensionnés (bande passante, déterminisme des couches sous-jacentes, etc.), de type LAN d'entreprise.