La voix simple en paquet n'est pas aussi contraignante que la parole téléphonique car elle n'implique aucune contrainte temporelle. Dans le cas d'IP, il ne faut donc pas confondre la téléphonie sur IP (ToIP) et la voix sur IP (VoIP).

Ce chapitre examine dans un premier temps l'évolution de la téléphonie vers les réseaux Internet et intranet puis aborde l'intégration téléphonie-informatique, aussi appelée CTI (Computer Telephony Integration).

L'application téléphonique

Comme expliqué précédemment, l'application de téléphonie est complexe à prendre en charge en raison de son caractère interactif et de sa forte synchronisation. Rappelons *(voir le chapitre 5)* les trois opérations successives nécessaires à la numérisation de la parole, qu'elle soit téléphonique ou non :

1. **Échantillonnage.** Consiste à prendre des points du signal analogique au fur et à mesure qu'il se déroule. Il est évident que plus la bande passante est importante, plus il faut prendre d'échantillons par seconde. C'est le théorème d'échantillonnage qui donne la solution : il faut échantillonner à une valeur égale à au moins deux fois la bande passante.
2. **Quantification.** Consiste à représenter un échantillon par une valeur numérique au moyen d'une loi de correspondance. Cette phase consiste à trouver la loi de correspondance de telle sorte que la valeur des signaux ait le plus de signification possible.
3. **Codage.** Consiste à donner une valeur numérique aux échantillons. Ce sont ces valeurs qui sont transportées dans le signal numérique.

La largeur de bande de la voix téléphonique analogique est de 3 200 Hz. Pour numériser ce signal correctement sans perte de qualité, puisqu'elle est déjà relativement mauvaise, il faut échantillonner au moins 6 400 fois par seconde. La normalisation a opté pour un échantillonnage de 8 000 fois par seconde. La quantification s'effectue par des lois semi-logarithmiques. L'amplitude maximale permise se trouve divisée en 128 échelons positifs pour la version américaine PCM, auxquels il faut ajouter 128 échelons négatifs dans la version européenne MIC. Le codage s'effectue donc soit sur 128 valeurs, soit sur 256 valeurs, ce qui demande en binaire 7 ou 8 bits de codage. La valeur totale du débit de la numérisation de la parole téléphonique s'obtient en multipliant le nombre d'échantillons par le nombre d'échelons, ce qui donne :

- 8 000 × 7 bit/s = 56 Kbit/s en Amérique du Nord et au Japon ;
- 8 000 × 8 bit/s = 64 Kbit/s en Europe.

Beaucoup d'autres solutions ont été développées par rapport aux qualités et aux défauts de l'oreille :

- AD-PCM (Adaptive Differential-Pulse Code Modulation), ou MIC-DA (Modulation par impulsion et codage-différentiel adaptatif) :
- SBC (Sub-Band Coding) ;
- LPC (Linear Predictive Coding) ;
- CELP (Code Excited Linear Prediction).

La section suivante fait un tour d'horizon des principaux codeurs audio.

Extrait : Livre "les Réseaux" G. Pujolle Ed. Eyrolles

Les codeurs audio

Les codeurs audio associés aux différentes techniques citées précédemment sont nombreux. On trouve notamment les codecs classiques mais aussi de nouveaux codeurs bas débit. La figure 35.1 illustre les vitesses de sortie des différentes normes de codeurs de la voix téléphonique fondées sur un échantillonnage standard à 8 kHz. L'ordonnée représente la qualité du son en réception, qui est évidemment un critère subjectif. Nous avons aussi représenté les codeurs utilisés dans les réseaux de mobiles GSM et les normes régionales.

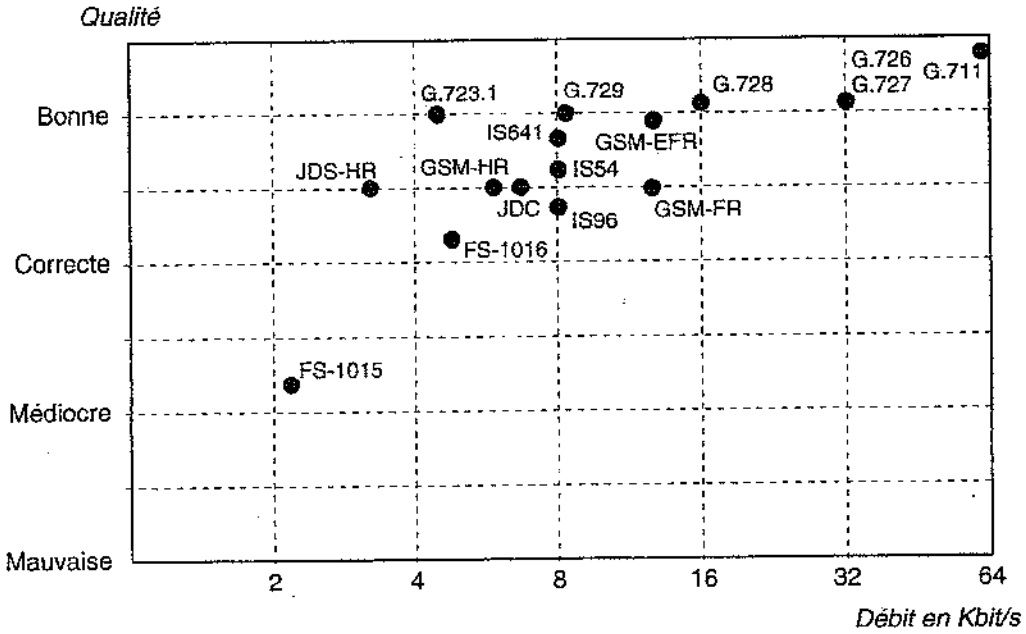

Figure 35.1
Codeurs audio

Pour l'audio haute définition, on considère une bande passante plus importante puisque l'oreille humaine est sensible aux fréquences de 20 à 20 000 Hz. L'échantillonnage s'effectue sur 40 kHz, et c'est la valeur de 44,1 kHz qui a été choisie. Le codage effectué sur un CD tient sur 16 bits par échantillon, ce qui donne 705,6 Kbit/s.

Parmi les nombreux codeurs propriétaires qui existent sur le marché, citons :

- StreamWorks à 8,5 Kbit/s ;
- VoxWare à 2,4 Kbit/s avec le codeur RT24 ;
- Microsoft à 5,3 Kbit/s avec la norme G.723 ;
- VocalTec à 7,7 Kbit/s.

La recommandation G.711 correspond à la numérisation classique à 64 Kbit/s en Europe ou 56 Kbit/s en Amérique du Nord. G.723 est une compression de la parole utilisée par de nombreux industriels, entre autres Microsoft, qui l'utilise dans l'environnement Windows. Le débit descend à presque 5 Kbit/s. G.726 est la norme adoptée pour la compression de la parole en codage différentiel adaptatif en 16, 24, 32 ou 40 Kbit/s.

Dans ce cas, au lieu de coder l'échantillon en entier, on n'envoie que la différence avec l'échantillon précédent, ce qui permet un codage sur beaucoup moins d'éléments binaires. G.727 utilise aussi un codage différentiel, qui apporte des compléments au codage précédent. Cette recommandation indique comment changer, en cours de numérisation, le nombre de bits utilisés pour coder les échantillons. Elle est particulièrement utile dans le cadre des réseaux qui demandent à l'application de s'adapter en fonction de la charge du réseau. G.728 est une compression à 16 Kbit/s utilisant une technique de prédiction, qui consiste à coder la différence entre la valeur réelle et une valeur estimée de l'échantillon à partir des échantillons précédents. On comprend que cette différence peut être encore plus petite que dans la technique différentielle. Si l'estimation est bonne, la valeur à transporter avoisine toujours 0. Très peu de bits sont alors nécessaires pour acheminer cette différence. Les standards FS proviennent du ministère américain de la Défense.

Les codeurs les plus récents sont G.723.1, G.729 et G.729.A. Le codeur G.723.1 permet un débit compris entre 5,3 et 6,4 Kbit/s. Les deux codeurs G.729 donnent un débit de 8 Kbit/s, mais la qualité de la communication est meilleure. Ce codec a été choisi pour compresser la voix dans l'UMTS.

La parole téléphonique est une application très contraignante, comme nous l'avons vu à plusieurs reprises dans cet ouvrage. La première contrainte provient de l'interactivité entre les deux utilisateurs, qui limite le temps aller-retour à une valeur de 600 ms au grand maximum. Les normes de l'UIT-T portent cette valeur à 800 ms. Cependant, pour avoir une bonne qualité de la communication, il faut descendre à 300 ms aller-retour. Suivant les protocoles sous-jacents, plusieurs méthodes permettant de satisfaire à ces contraintes ont été développées à la fin des années 90, que nous allons examiner.

La téléphonie sur ATM et le relais de trames

La technique de transfert ATM a été conçue pour transporter de la parole téléphonique de type G.711 à 64 Kbit/s. La raison de la petite taille de la cellule se trouve dans cette fonctionnalité. Les 48 octets de données de la trame sont remplis en 48 fois 125 µs, c'est-à-dire 6 ms, ce qui reste acceptable, même lorsqu'il y a des échos et que le temps de transit doit rester inférieur à 28 ms. Si la parole téléphonique est compressée par un codeur G.729 à 8 Kbit/s, il faut un temps de 48 ms de remplissage des 48 octets de données puisque le signal donne naissance à 1 octet toutes les 1 ms. Cette section examine la technique AAL-2 introduite dans la commutation ATM pour réaliser le transport de la voix téléphonique et plus particulièrement la téléphonie UMTS. Avant d'aborder l'AAL-2, introduisons les techniques préalables, qui sont encore utilisées dans les réseaux ATM.

L'émulation de circuit CES (Circuit Emulation Service) a été la première solution pour transporter de la téléphonie en paquet. Cette émulation de circuit utilise l'AAL-1 de l'environnement ATM, et plus précisément le service CBR (Constant Bit Rate), présenté au chapitre 15. Les PABX interconnectés par cette solution utilisent des interfaces E1 normalisées (G.703 et G.704). Le service ATM est de type circuit virtuel permanent. La signalisation sur l'interface est portée dans l'IT16 de l'interface E1.

La téléphonie sur IP

La parole sur les réseaux IP a été abordée au chapitre 1 lors de l'examen du transport d'applications isochrones. La problématique du transport de la parole téléphonique dans des environnements IP est assez différente suivant que l'on est sur un réseau IP non contrôlé, comme Internet, ou sur un réseau permettant l'introduction d'un contrôle, comme le réseau privé d'une compagnie, de type intranet, ou celui d'un ISP.

Sur l'Internet de première génération, il faut que le réseau soit peu chargé pour que la contrainte de 300 ms soit respectée. Sur les réseaux intranet et ceux des fournisseurs d'accès à Internet, mais aussi ceux des opérateurs, le passage de la parole est possible à condition de contrôler le réseau pour que le temps total de transport, y compris la paquétisation et la dépaquétisation, soit limité.

De nombreuses solutions ont été proposées, comme VoIP (Voice over IP) de l'IMTC (International Multimedia Teleconferencing Consortium). Dans ces solutions, il a d'abord fallu définir un codeur normalisé. Le choix s'est généralement porté sur G.723, mais d'autres solutions sont opérationnelles, comme le codeur G.711. Le paquet IP doit être le plus court possible, et il faut multiplexer plusieurs voies de parole dans un même paquet, de façon à raccourcir le temps de remplissage et à limiter les temps de transfert dans le réseau. Si les routeurs peuvent gérer des priorités, ce qui est possible en utilisant des services de type DiffServ, la parole téléphonique est acheminée beaucoup plus facilement dans le laps de temps demandé.

Plusieurs organismes de normalisation, de droit ou de fait, travaillent sur ce sujet particulièrement prometteur. Dans les organismes de droit, l'ETSI, l'organisme de normalisation européen, a mis sur pied le groupe TIPHON (Telecommunications and Internet Protocol Harmonization Over Networks). Le projet porte sur la parole et le fax entre utilisateurs connectés, en particulier sur des réseaux IP. Le cas où un utilisateur travaille sur un réseau IP et un autre sur un réseau à commutation de circuits, qu'il soit téléphonique, RNIS, GSM ou UMTS, entre également dans le cadre des études de TIPHON. Les activités de TIPHON concernent en outre la validation de solutions pour transporter la parole téléphonique par le biais de démonstrateurs. Il s'agit d'expériences en vraie grandeur visant à démontrer l'efficacité des solutions. L'ETSI travaille pour cela en collaboration avec l'UIT-T et l'IETF mais aussi avec les groupes IMTC et VoIP.

L'UIT-T travaille de son côté activement sur le problème de la téléphonie sur IP dans trois groupes du SG 16 (voir en annexe pour le détail des groupes de travail de l'UIT-T) : le WP1 pour les modems (série V), le WP2 pour les codecs (série G) et le WP3 pour les terminaux (série H). L'objectif de l'UIT-T est de développer un environnement complet et non pas simplement un terminal ou un protocole.

Au sein de l'IETF, de nombreux groupes de travail s'attaquent à des problèmes spécifiques, parmi lesquels :

- AVT (Audio Video Transport), qui utilise le protocole RTP (RFC 1889 et 1890) pour effectuer la communication en temps réel.
- MMUSIC (Multiparty Multimedia Session Control), qui utilise le protocole SIP, présenté en détail au chapitre 31 et que nous explicitons à nouveau un peu plus loin dans ce chapitre.

- IPTel (IP Telephony), qui définit un protocole de localisation des passerelles et un langage permettant de mettre en communication des circuits et des flots IP.
- PINT (PSTN IP Internetworking), qui utilise également le protocole SIP.
- FAX (Fax over IP), qui stocke et émet des fax par l'intermédiaire de messages électroniques.
- MEGACO (Media Gateway Control), qui détermine un protocole entre une passerelle et son contrôleur.
- SIGTRAN (Signal Translation), qui propose l'utilisation du passage des commandes de la signalisation CCITT n° 7 dans des paquets IP.
- ENUM (E.164/IP translations), qui gère les translations d'adresses E.164 vers des adresses IP.

Respecter la contrainte temporelle est une première priorité pour le transport de la parole téléphonique. Une seconde priorité concerne la mise en place d'une signalisation pour mettre en connexion les deux utilisateurs qui veulent se parler.

Les protocoles de signalisation utilisés pour le transport et la gestion de la parole sous forme de paquets IP regroupent essentiellement H.323 et SIP (Session Initiation Protocol). Nous verrons en détail le protocole H.323 au chapitre suivant. Ce protocole a été défini dans un environnement de télécommunications, à la différence de SIP, qui provient de l'informatique et plus spécifiquement du Web. SIP peut utiliser le code HTTP ainsi que la sécurité qui y est liée. Il peut en outre s'accommoder des pare-feu de protection. SIP met en place des sessions, qui ne sont que des appels téléphoniques entre un client et un serveur. Six primitives HTTPsont utilisées pour cela : INVITE, BYE, OPTIONS, ACK, REGISTER et CANCEL.

La VoIP devient peu à peu une application classique grâce aux possibilités de numérisation et à la puissance des PC, qui permettent d'annuler les échos. L'élément le plus contraignant reste le délai, surtout lorsqu'il faut traverser des terminaux de type PC, des modems, des réseaux d'accès, des passerelles, des routeurs, etc.

On peut considérer que le PC demande un temps de traversée d'une centaine de millisecondes, le modem de quelques dizaines de millisecondes, la passerelle également d'une centaine de millisecondes et le réseau IP de quelques dizaines de millisecondes. Le total montre que la limite des 300 ms pour avoir une interactivité est rapidement atteinte. Si l'on dépasse les 150 ms de transit et que l'on s'approche des 300 ms, la qualité de la communication s'en ressent, comme lors d'une conversation par satellite.

Détaillons la mise en place de la communication. Il faut utiliser une signalisation pour mettre en place la session. Premier élément, la localisation du récepteur (User Location) s'effectue par une mise en correspondance de l'adresse du destinataire (adresse IP ou téléphonique classique) en une adresse IP. Le protocole DHCP et les passerelles spécialisées sont des éléments de solution pour déterminer les adresses des récepteurs. L'établissement de la communication passe par une acceptation du terminal destinataire, que ce soit un téléphone, une boîte vocale ou un serveur Web. Comme nous l'avons vu, plusieurs protocoles de signalisation peuvent être utilisés, comme H.323, de l'UIT-T, SIP ou SDP, de l'IETF.

Les protocoles de signalisation

Nous avons déjà examiné en détail le protocole SIP au chapitre 31. Rappelons quelques éléments de ce protocole avant d'aborder SDP et surtout RTP-RTCP.

Comme son nom l'indique, SIP (Session Initiation Protocol) est utilisé pour initialiser la session. Une requête SIP contient un ensemble d'en-têtes qui décrivent l'appel, suivis du corps du message, contenant la description de la demande de session. SIP est un protocole client-serveur, qui utilise la syntaxe et la sémantique de HTTP. Le serveur gère la demande et fournit une réponse au client.

Trois types de serveurs gèrent différents éléments : un serveur d'enregistrement (Registration Server), un serveur relais (Proxy Server) et un serveur de redirection (Redirect Server). Ces serveurs travaillent à trouver la route. Le serveur proxy détermine le prochain serveur (Next-Hop Server), qui, lui-même, trouve le suivant, et ainsi de suite. Des champs supplémentaires de l'en-tête précisent les options, comme le transfert d'appel ou la gestion de conférence téléphonique.

Le protocole SDP (Session Description Protocol) est utilisé pour décrire les sessions multimédias pour la partie téléphonique mais aussi pour d'autres applications distribuées, comme la radio sur Internet.

SDP permet le transfert de nombreuses informations, notamment les suivantes :

- flots correspondant aux médias de l'application ;
- pour chaque flot, adresse de destination, unicast ou multicast ;
- pour chaque flot, numéro de port UDP ;
- type de charge transportée ;
- instants de synchronisation (par exemple, l'instant de début d'un programme de télévision diffusée) ;
- origine de la demande de communication.

Le protocole RTP (Real-time Transport Protocol) prend le relais pour le transport de l'information proprement dite *(voir le chapitre 18)*. Son rôle est d'organiser les paquets à l'entrée du réseau et de les contrôler à la sortie pour reformer le flot avec ses caractéristiques (synchronisme, perte, etc.). C'est un protocole qui travaille au niveau transport et essaye de corriger les défauts apportés par le réseau.

Les fonctions de RTP sont les suivantes :

- Le séquencement des paquets par une numérotation permettant de détecter les paquets perdus, ce qui est essentiel pour la recomposition de la parole. La perte d'un paquet n'est pas en soi un problème, s'il n'y en a pas trop de perdus. En revanche, repérer qu'un paquet a été perdu est impératif car il faut en tenir compte et éventuellement le remplacer par une synthèse déterminée en fonction des paquets précédant et suivant.
- L'identification de ce qui est transporté dans le message pour permettre, par exemple, une compensation en cas de perte.
- La synchronisation entre médias, grâce à des estampilles.

- L'indication de tramage. Les applications audio et vidéo sont transportées dans des trames dont la dimension dépend des codecs effectuant la numérisation. Ces trames sont incluses dans les paquets pour être transportées et doivent être récupérées facilement au moment de la dépaquétisation afin que l'application soit décodée simplement.
- L'identification de la source. Dans les applications en multicast, l'identité de la source doit être déterminée.

RTP utilise le protocole RTCP (Real-Time Control Protocol) pour transporter les informations supplémentaires suivantes pour la gestion de la session :

- Retour de la qualité de service lors de la demande de session. Les récepteurs utilisent RTCP pour renvoyer vers les émetteurs des rapports sur la QoS. Ces rapports comprennent le nombre de paquets perdus, la gigue et le délai aller-retour. Ces informations permettent à la source de s'adapter, c'est-à-dire, par exemple, de modifier le degré de compression pour maintenir la QoS.
- Synchronisation supplémentaire entre médias. Les applications multimédias sont souvent transportées par des flots distincts. Par exemple, la voix et l'image, ou même une application numérisée sur plusieurs niveaux hiérarchiques, peuvent voir les flots générés suivre des chemins distincts.
- Identification. Les paquets RTCP contiennent des informations d'adresse, comme l'adresse d'un message électronique, un numéro de téléphone ou le nom d'un participant à une conférence téléphonique.
- Contrôle de la session. RTCP permet aux participants d'indiquer leur départ d'une conférence téléphonique (paquet Bye de RTCP) ou simplement une indication de leur comportement.

Le protocole RTCP demande aux participants de la session d'envoyer périodiquement ces informations. La périodicité est calculée en fonction du nombre de participants à l'application.

Un autre protocole utilisable est RTSP (Real-Time Streaming Protocol), dont le rôle est de contrôler une communication entre deux serveurs où sont stockées des informations multimédias audio et vidéo. RTSP offre des commandes assez semblables à celles d'un magnétoscope, telles que avance, avance rapide, retour, pause, etc. Ce protocole peut être très utile dans le cadre de la téléphonie sur IP en permettant l'enregistrement d'une téléconférence pour la réentendre ultérieurement, la vision d'une séquence vidéo, l'enregistrement de message téléphonique, etc.

Un autre point important pour réaliser la communication de l'émetteur vers le récepteur concerne les fonctionnalités de la passerelle permettant de passer d'un réseau à transfert de paquets à un réseau à commutation de circuits, avec les problèmes d'adressage, de signalisation et de transcodage que cela pose. Ces passerelles se démultiplient entre ISP et opérateurs télécoms.

Pour finaliser l'ouverture d'un appel, le protocole SIP envoie une requête à la passerelle. Le premier problème est de déterminer quelle passerelle est capable de réaliser la liaison circuit pour atteindre le destinataire. En théorie, chaque passerelle peut appeler n'importe quel numéro de téléphone. Cependant, pour réduire les coûts, il vaut mieux choisir une passerelle locale.