

Visioconférence : les technologies d'aujourd'hui

Sommaire

Deux modes de transmission :

- Transmission via Numeris
- Transmission sous IP. Les protocoles

Deux familles de normes : H320 et H323

- Des normes et des formats
- Transmettre la vidéo et l'audio
- Présenter et partager des documents
- Nouveautés et évolutions

Les contraintes liées aux réseaux locaux et à leurs équipements spécifiques

- La gestion des adresses IP
- Le passage des firewalls
- Un outil particulier : le Gatekeeper

La visioconférence en multi points

Interconnecter les deux familles : les passerelles

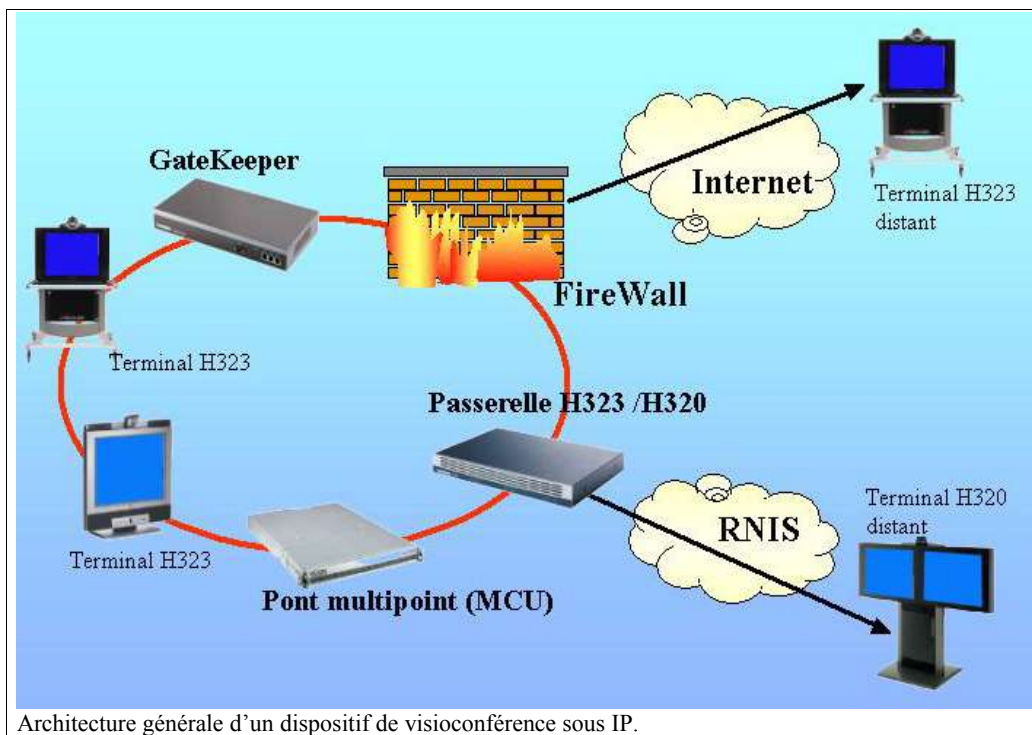
On appelle visioconférence la technologie multimédia qui permet à deux ou plusieurs interlocuteurs distants de communiquer en temps réel par l'intermédiaire d'un dispositif de sonorisation et de visualisation (écrans, téléviseurs) et de travailler sur des documents communs (textes imprimés, photographies, médias audiovisuels ou informatiques...).

La mise en relation de deux groupes de personnes grâce à l'image et au son s'appuie sur deux entités technologiques distinctes : d'une part, un réseau de transmission et d'autre part, des équipements et des terminaux spécifiques pour les utilisateurs. Des dispositifs particuliers ont été développés pour permettre des relations entre plusieurs groupes (liaison multipoint).

Le réseau de transmission assure l'acheminement de toutes les informations (images, sons, données informatiques, signaux de services...) entre les différents points. Traditionnellement, cette fonction s'opérait via le Réseau Numérique de France Telecom plus connu sous le nom commercial de Numeris. Avec le développement de l'Internet et des

réseaux d'entreprises, on assiste aujourd'hui à une véritable montée en puissance de dispositifs qui s'appuient désormais sur les technologies et sur les protocoles utilisés pour Internet (et plus communément nommés « sous IP » pour Internet Protocol).

Les équipements pour les utilisateurs sont placés aux points terminaux. Qu'elle que soit la technologie utilisée pour l'acheminement des données, ces équipements vont reposer sur les mêmes principes de base, même s'ils ne s'appuient pas exactement sur les mêmes normes. Un dispositif de visioconférence, qu'il soit construit autour d'un micro-ordinateur pour un usage individuel ou autour d'équipements spécifiques pour une utilisation collective, sera toujours composé de la même série de composants : une (ou plusieurs) caméra, un (ou plusieurs) microphone, les codecs, les dispositifs de visualisation (téléviseurs, moniteurs, vidéo projecteur, écrans numériques), l'interface réseau... Des éléments annexes (pont, passerelle, gatekeeper...) pourront être ajoutés le cas échéant lorsque des fonctionnalités plus étendues seront nécessaires.



Ce document fera le point sur toutes ces technologies. Cependant, compte tenu du fort développement des visioconférences sous IP, l'accent sera donné sur cette technologie.

Le dernier chapitre abordera la problématique des réseaux locaux et des contraintes qu'ils induisent généralement dans le bon fonctionnement d'une session de visioconférence.

Cette étude se limitera aux utilisations traditionnelles, (en entreprise, dans les lycées...) mais n'abordera pas la problématique des établissements de l'enseignement supérieur qui, en ayant la possibilité d'être raccordés au réseau à haut débit Renater, disposent de spécificités et d'outils particuliers qui ne sont pas encore proposés sur «l'Internet classique».

Deux modes de transmission

Pour établir une visioconférence entre deux sites, deux modes de transmission sont donc aujourd'hui privilégiés : le réseau Numeris de France Telecom d'une part, les réseaux IP au sens large du mot, c'est à dire le réseau Internet dans son aspect mondial, et les réseaux informatiques locaux (de type Ethernet par exemple...) qui y sont connectés. Les deux modes de transmission ont en commun d'être tous les deux entièrement numériques, mais la similitude s'arrête là. Des différences fondamentales existent. Les deux modes de diffusion sont technologiquement très différents et s'appuient sur deux familles de normes spécifiques : l'un est plus ancien, et bien ancré dans les habitudes, le second est plus récent et se développe actuellement de manière importante même s'il n'est pas toujours le plus performant.

Les liaisons entre les sites distants sont en temps réel et en full duplex (les liaisons sont bi-directionnelles et chaque site est simultanément émetteur et récepteur). Les débits sont identiques dans les deux sens.

La majorité des matériels commercialisés en France depuis 5 ans peuvent fonctionner sous IP. Un grand nombre d'entre eux adoptent la double compatibilité IP et RNIS.

Numeris

Le réseau de télécommunication numérique de France Telecom (RNIS pour Réseau Numérique à Intégration de Service, en anglais ISDN - Integrated Service Digital Network), plus connu en France sous le nom commercial de NUMERIS, est disponible en France depuis le début des années 1990. C'est un réseau numérique de bout en bout (par opposition au réseau téléphonique (RTC) où les liaisons terminales avec l'utilisateur sont toujours analogiques) et le débit y est garanti. Un accès dit de base comprend deux canaux B à 64 kb/s pour les données soit un débit utile de 128 Kb/s, et un canal pour la signalisation (canal D à 16 Kb/s). Des débits supérieurs, 256 Kb/s ou 384 Kb/s, sont possibles par le regroupement de plusieurs accès de base. Des couplages plus importants sont également réalisables : un accès dit « primaire » comprend 30 canaux B à 64 Kb/s ce qui correspond alors à un débit utile de 2 Mb/s. La qualité des images (et dans une moindre mesure des sons) est directement dépendante du débit possible dans le réseau de transmission : plus il sera élevé et meilleure sera la restitution. Pour obtenir une fluidité correcte des images, il est nécessaire d'opter au minimum pour une liaison à 384 Kb/s.

A l'image d'une communication téléphonique classique, on établit une connexion temporaire pour mettre en relation les interlocuteurs distants (on utilise le terme de commutation de circuit). Elle permet l'établissement d'un chemin direct qui est dédié à la communication entre les points participants. Toutes les données vont suivre le même chemin pendant toute la durée de la session. La qualité de service est assurée : le débit étant garanti, la qualité des images et des sons n'est pas sujette à variations, voire à des coupures. On a une qualité de service optimum, ce qui n'est pas le cas, on le verra avec les technologies sous IP.

Numeris est disponible sur abonnement auprès de France Télécom. Le RNIS est pratiquement disponible dans le monde entier et est normalisé. Comme pour le téléphone, les communications sont facturées en fonction de la durée de la communication et de la distance. Le coût est également proportionnel au nombre de canaux B utilisés. De ce fait, dans la majorité des cas, les débits adoptés ne dépasseront pas 384 Kb/s.

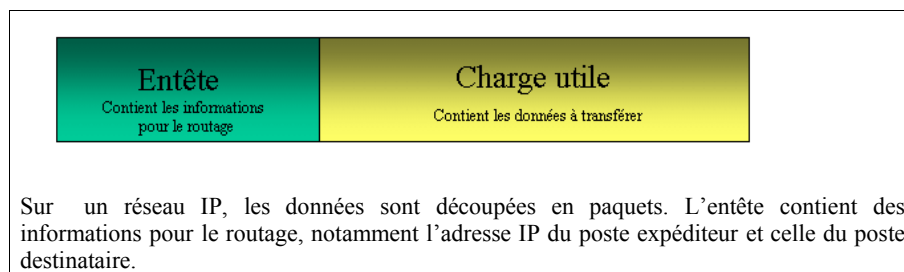
La visioconférence sous IP : des avantages mais aussi des contraintes spécifiques

Le titre générique de « visioconférence sous IP » est utilisé dans son sens le plus large, c'est à dire pour désigner des visioconférences qui s'appuient à la fois sur les technologies d'Internet mais aussi sur celles des réseaux privés ou locaux (réseaux Ethernet, Intranet ...).

A l'inverse de Numeris, la visioconférence sur IP repose, en ce qui concerne sa transmission sur les réseaux, sur des principes techniques totalement différents. Cette technologie apporte de nombreux avantages mais engendre aussi des inconvénients spécifiques.

La visioconférence sur IP utilise une infrastructure qui n'a pas été conçu à l'origine pour des applications vidéo. Que ce soit sur Internet ou dans le cadre du réseau informatique d'un établissement, les lignes utilisées sont parcourues par des flux divers et variés, leurs capacités en terme de bande passante doivent être partagées en permanence entre de nombreuses applications ou périphériques informatiques... Le débit possible (et donc la qualité de la transmission) dépend directement de la charge du réseau à un moment donné et peut donc être très variable en fonction du moment. Le débit n'y est pas garanti d'où le risque d'une mauvaise fluidité des images et des sons.

A l'opposé de la visioconférence sur RNIS pour laquelle on établit une liaison directe et permanente entre les participants et où toutes les informations numériques circulent sous la forme d'un flux continu avec un débit constant, (et donc avec une qualité de service constante), les données numériques sur un réseau IP sont découpées en paquets. Ces paquets circulent sur le réseau et sur Internet totalement indépendamment les uns des autres. Ils peuvent, en toute autonomie, prendre des itinéraires différents (en plus des données « utiles », chacun d'entre eux intègre une entête contenant différentes informations, notamment l'adresse IP d'arrivée). Ces paquets peuvent parvenir à destination avec des écarts variables et pas nécessairement dans le bon ordre voire être détruits en cours de route. De prime abord, ces caractéristiques pourraient paraître totalement incompatibles pour une restitution correcte de la vidéo et de l'audio qui exige par nature un flux continu et régulier. Sur le plan qualitatif, les problèmes de transit évoqués peuvent se matérialiser à l'écran par une dégradation des images (apparition d'artefacts) voire leur gel ou leur disparition complète. Des technologies particulières ont été développées pour adapter les réseaux IP au contraintes du transport des données vidéo et audio en temps réel.



Conçu à l'origine pour des applications qui n'étaient pas multimédia, Internet repose à la base sur deux protocoles IP et TCP. Le protocole IP assure l'acheminement des paquets de point en point, jusqu'au terminal final mais sans se préoccuper du contenu. Il ne gère pas les pertes et les retards. Ce protocole simple mais peu fiable est complété par le protocole TCP qui assure la fiabilité de la transmission en demandant la ré-émission des paquets perdus ou détruits. Du fait de cette procédure de ré-émission, TCP est un protocole lent. Cette fiabilité

qui est un atout pour la transmission de fichiers « informatiques » devient, du fait de la lenteur, un inconvénient pour la transmission de la vidéo et de l'audio.

Encadré 1 : Dans le cadre d'une visioconférence, les échanges sont symétriques. Sur un réseau IP, le débit pour chacune des directions est la somme des débits nécessaires pour l'audio, la vidéo, et les données. Il faut y ajouter les informations nécessaires à la transmission des paquets et qui sont contenues dans l'entête de chacun des paquets (compter environ 20%). Une visioconférence établie entre deux centres avec une vidéo à 384 Kb/s et de l'audio à 64 Kb/s, engendrera sur le réseau un débit total de deux fois $(384 + 64) \times 1,2 = 540$ Kb/s soit un total de plus de 1 Mb/s.

Encadré 2 : Un des principaux problèmes des réseaux est la congestion ou la surcharge du trafic qui va provoquer des retards et les délais dans l'acheminement des paquets (à l'arrivée, les écarts temporels entre paquets ne sont pas identiques à ce qu'ils étaient au départ : on appelle cela la gigue). On utilise souvent le terme de "qualité de service" (QoS) pour caractériser l'aptitude d'un réseau à assurer la transmission des données.

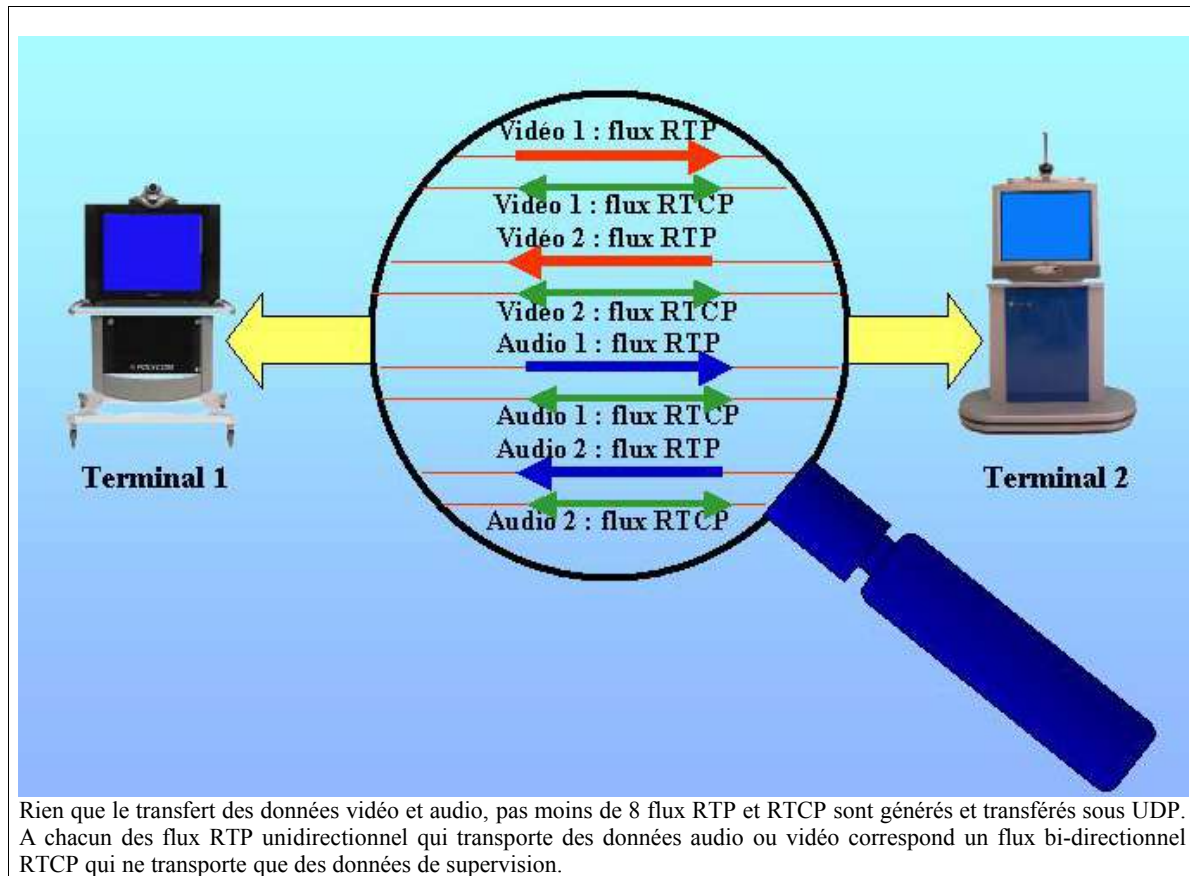
La qualité de service "best effort" est celle qui est fournie traditionnellement par l'Internet et correspond en fait à une absence de qualité de service. La notion de priorité n'existe pas, le réseau traite de façon identique tous les flux qui circulent à chaque instant, quel qu'en soit le contenu. Les dégradations (ralentissements, pertes) dues aux éventuelles surcharges sont réparties à égalité entre tous. Le réseau "fait de son mieux" (best effort) pour acheminer le plus efficacement possible la totalité des données, sans s'inquiéter des besoins spécifiques correspondant à chacun des flux, et donc sans établir de différenciation entre les données audiovisuelles et les données « informatiques ».

Pour compenser la lenteur de TCP, on a ainsi créé le protocole UDP (User Datagram Protocol), un protocole simplifié à l'extrême, sans ré-émission des paquets perdus mais qui présente l'avantage d'être beaucoup plus rapide. Au cours d'une session de visioconférence, les deux protocoles TCP et UDP sont simultanément utilisés conjointement avec IP en fonction des caractéristiques des données à transmettre : TCP pour les données ne souffrant aucune perte (établissement des appels, signalisation et gestion des communications, applications informatiques partagées...) et UDP pour la transmission bilatérale des signaux audios et vidéos pour lesquels une transmission en temps réel est exigée et où des pertes peuvent être tolérées.

Ces protocoles majeurs ont été complétés par des protocoles spécifiques et par des mécanismes particuliers destinés à optimiser la transmission des données audiovisuelles.

Le protocole RTP (Real-time Transport Protocol) contrôle les flux vidéo et audio dans les applications en temps réel. Il assure la numérotation des séquences, ajoute une référence temporelle (timestamp) qui indique l'instant exact d'émission du paquet à la source permettant ainsi à l'arrivée de replacer les paquets dans le bon ordre, et de rétablir la régularité temporelle. Sous le terme de Packet Assist, la société VCON rassemble sur ses matériels tout un package de fonctionnalités destinées justement à compenser tous ces défauts temporels : Packet Ordering pour replacer les paquets dans le bon ordre, Jitter Correction pour recalibrer leur séquençement, Lip Sync Correction pour réaliser la re-synchronisation des données audio et vidéo et Lip Sync Delay Adjustment pour modifier éventuellement le décalage entre le son et l'image (et compenser une désynchronisation

possible entre le mouvement des lèvres et la voix des intervenants). Des dispositifs similaires sont proposés chez les autres fournisseurs.



L'optimisation de la qualité des transmissions passe également par l'implémentation de mécanismes spécifiques au niveau des applications terminales et des éléments intermédiaires du réseau (les routeurs).

Une première étape est d'adapter en temps réel le débit de la vidéo en fonction des capacités instantanées du réseau (et donc d'assurer une qualité des images aussi optimale que possible). Cette adaptation repose sur les échanges qui sont établis à intervalles réguliers entre les organes terminaux et qui s'appuient sur les protocoles RTP et RTCP (Real-time Transport Control Protocol). En fonction des « compte-rendus » de réception émis par le terminal, la station émettrice modifie les paramètres de compression et de diffusion des données vidéo (et parfois aussi audio). La qualité de la restitution va ainsi diminuer légèrement lorsque le débit sur le réseau devient plus faible (pour les images vidéo, la définition sera un peu moins bonne, la fluidité moins soutenue, pour l'audio, la bande passante sera plus réduite ...) et vice et versa. Dans un réseau provisoirement congestionné, il est sans doute préférable d'afficher des images avec une qualité amoindrie - plus fortement compressées, elles nécessitent un débit plus faible - mais partiellement exemptes de défauts de transmission (peu de pertes de paquets, jitter faible) que des images de meilleure facture (compressées pour un débit plus élevé) mais qui présenteraient, compte tenu de la saturation du réseau, des taux de pertes et de jitter importants et donc de nombreux artefacts, voire un gel complet.

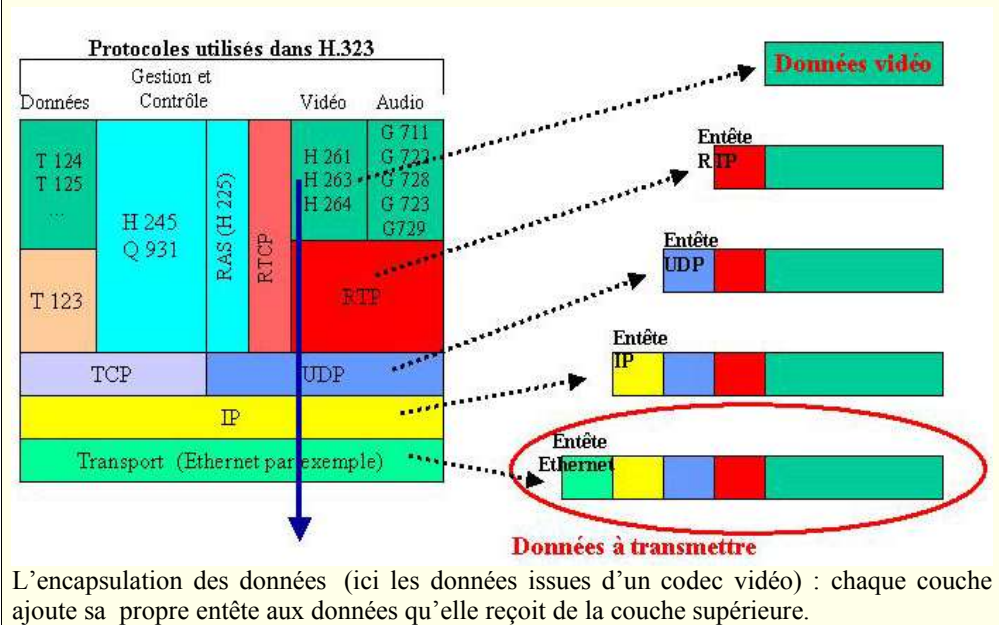
Encadré : Sur un réseau IP, la circulation des données obéit à trois grands principes.

Avec mon premier, les données utiles ne circulent pas en un flot continu sur le réseau mais sont « découpées » en paquets. A chacun d'entre eux est affecté une entête (header en anglais) qui contient des informations pour le service.

Avec mon second, le traitement de l'information est hiérarchisé sous la forme de couches fonctionnelles, chacun d'entre elles correspondant à une fonctionnalité précise (couche application, couche transport, couche réseau...) et donc à un ou plusieurs protocoles spécifiques.

Concrètement, le flux d'informations transformé en paquets est transmis de couche en couche. Chacune d'entre elles ajoute cependant ses propres informations de service aux données utilisateurs correspondant aux divers protocoles utilisés. A la manière de poupées russes s'emboîtant les unes dans les autres, chaque paquet avec sa charge utile et son entête devient la charge utile de la couche suivante et est complétée par une entête (et aussi, le cas échéant, par des octets de fin). On parle d'encapsulation, c'est le troisième principe.

La succession de nombreux protocoles, si elle répond à la nécessité d'améliorer la transmission des données, a aussi pour conséquence négative un accroissement important des données « de service » comparativement au volume des informations « utiles » et donc une augmentation du volume total à transmettre. Ces entêtes successives contiennent des informations qui sont, pour certaines, identiques d'une entête à l'autre. Des techniques de compression ont été élaborées pour alléger le volume de ces entêtes et permettre la suppression de ces redondances à l'émission.



Les deux protocoles RTP et RTCP n'ont donc pas pour mission d'agir sur les équipements constitutifs du réseau. Dans cette quête pour l'optimisation des transmissions, une deuxième étape est d'intervenir sur tous les éléments intermédiaires. Le transfert des paquets sur le réseau peut être amélioré,

- soit en « balisant » le chemin et en réservant de la bande passante : la réservation de ressources s'établit via le protocole de réservation RSVP (Ressource Reservation Protocol). Ces réservations sont demandées par le récepteur et adressées périodiquement aux différents routeurs.
- soit en affectant des priorités aux paquets contenant des données sensibles : les données

audio et vidéo sont rendues prioritaires par rapport aux autres données lorsque des congestions se produisent sur des routeurs.

Ces deux derniers points, s'ils peuvent être mis en oeuvre sur des réseaux privés ou sur des réseaux universitaires (Renater), restent encore pour aujourd'hui largement inexploités sur Internet.

La qualité d'une visioconférence est également tributaire des caractéristiques du réseau informatique local. Sur un réseau de type Ethernet, les débits sont élevés, dans la majorité des cas 100 Mb/s (les réseaux à 10 Mb/s sont à éviter). Même si la bande passante disponible est à partager entre tous les utilisateurs du réseau (c'est avant tout un réseau informatique), elle est généralement suffisante pour assurer le transit des données nécessaires à une visioconférence. Les réseaux Ethernet commutés sont préférables à des réseaux partagés, car ils offrent de meilleures performances en terme d'utilisation de la bande passante. En scrutant l'adresse du destinataire dans chaque paquet IP, le switch ou commutateur n'aiguille le paquet que vers la sortie correspondant à la seule machine destinataire, par opposition au HUB avec lequel un message émis par une machine est « visible » par l'ensemble des machines, même s'il ne peut être exploité que par celle qui est destinataire. Le switch est un aiguillage automatique, le hub une prise multiple !

La fluidité dans le réseau pourra être également améliorée par la présence d'un équipement particulier, le GateKeeper. Ce dispositif spécifique pour la visioconférence (mais facultatif) a principalement un rôle d'annuaire, mais il peut également participer à la gestion des flux en régulant le trafic supplémentaire généré par la séance de visioconférence : il pourra par exemple interdire à un interlocuteur de participer à une session si le trafic généré risque de saturer le réseau. Ce point sera abordé dans un prochain paragraphe.

Dans un établissement scolaire câblé, en s'appuyant sur un réseau qui est généralement largement déployé dans la quasi totalité des salles, il est possible de réaliser facilement et rapidement une session de visioconférence en tout lieu, puisqu'il suffit simplement de disposer d'une prise « réseau » à proximité pour y connecter un équipement, qu'il soit individuel ou destiné à un usage collectif. Cette souplesse d'utilisation n'est pas possible avec le RNIS, les lignes n'étant généralement déployées que dans quelques salles spécifiques. A ces facilités, s'ajoutent des avantages économiques : pas de coût de communication, pas d'abonnement spécifique, si ce n'est celui de la connexion à haut débit à Internet, mais déjà comptabilisé par ailleurs car indispensable. A noter que sur les accès de type ADSL qui sont majoritairement proposés aujourd'hui, les liaisons sont asymétriques : les débits sont différents dans le sens montant et dans le sens descendants, par exemple 1024 Kb/s en réception mais seulement 128 Kb/s en émission¹. Dans une session de visioconférence simple entre deux points, les débits nécessaires sont identiques dans les deux sens (les liaisons sont symétriques). Dans l'exemple précédent, sauf fonctionnalité spécifique, c'est la valeur commune (donc ici 128 Kb/s) qui pourrait être automatiquement adoptée par les deux équipements.

¹ La technologie DSL (Digital Subscriber Line) est une grande famille de normes. L'ADSL (Asymmetric Digital Subscriber Line), asymétrique comme son nom l'indique, est l'un de ses éléments. Des versions symétriques existent aussi.

Deux familles de normes : H320 pour RNIS et H323 pour IP

Des normes et des formats

Pour ces deux méthodes de transmission, des normes spécifiques ont été établies afin de garantir l'interopérabilité de tous les matériels de visioconférence. Elaborée au début des années 90, la norme H320 a été prévue pour les réseaux à commutation de circuits et s'applique donc aux communications sur les lignes RNIS. La norme H323 concerne les réseaux à commutation de paquets, c'est à dire notamment aux réseaux IP. Elles ont été développées par l'ITU (International Telecommunication Union).

Les deux normes H320 et H323 sont des normes « conteneurs » et correspondent en fait à un assemblage de normes spécifiques pour tous les domaines concernés. Ces dernières peuvent être identiques pour H320 et H323 ou bien différentes. Elles concernent l'établissement et la gestion des communications, le contrôle et la signalisation (H225, Q931 pour l'initialisation des appels, H245 pour la négociation des canaux des médias...), le traitement de la vidéo et de l'audio, le partage d'applications (T120), le contrôle des caméras distantes (H281 en mode RNIS, H282 et H283 en mode IP...) ou le fonctionnement en multipoint (H243). Ces normes garantissent le bon fonctionnement des différentes phases d'une visioconférence et l'interopérabilité entre des matériels d'origines différentes.

Pour la transmission des données sur un réseau IP, la norme H323 s'appuie par ailleurs sur les protocoles de transport (TCP, UDP, RTP...) édictés par l'IETF (Internet Engineering Task Force) et évoqués dans le chapitre précédent.

Transmettre la vidéo et l'audio

Les images vidéo exigent des débits importants comparativement aux autres médias. Des technologies spécifiques ont été développées afin de diminuer la quantité d'informations devant être transmise et par la même diminuer le débit nécessaire et le rendre compatible avec les capacités des lignes de transmission utilisées (RNIS ou IP)². Cette réduction s'accompagne inévitablement d'une altération de la qualité des images et de leur fréquence de défilement. On appelle CODEC l'entité chargée de la Compression des données audio ou vidéo dans un sens (à la prise de vue) et de leur DECompression dans le sens contraire (pour en permettre l'affichage). Plus le débit est élevé et meilleure sera la qualité de la restitution, l'idéal étant d'obtenir la meilleure qualité possible pour un débit qui reste le plus faible (c'est à dire un taux de compression le plus élevé possible).

² Quelques chiffres pour fixer les esprits : une séquence vidéo numérique non compressée de qualité studio de télévision exige un débit de 166 Mb/s. La diffusion d'un programme télévisé dans le cadre d'un bouquet numérique satellitaire (TPS, Canal Sat) ou par un réseau câblé urbain s'effectue avec des débits de 4 à 6 Mb/s. Pour les matériels de visioconférence, les débits proposés sur les équipements sont au maximum de 1 à 2 Mb/s, mais restent le plus souvent limités à quelques centaines de Kb/s.

Compresser une séquence vidéo, c'est diminuer la quantité d'informations la caractérisant.

C'est donc réduire

- ☐ la dimension des images,
- ☐ leur fréquence d'affichage (avec pour corollaire une diminution de la fluidité !)

mais c'est aussi exploiter

- ☐ les redondances spatiales (par exemple, les plages uniformes à l'intérieur de chacune des images),
- ☐ les redondances temporelles (dans une séquence vidéo, les différences entre deux images successives sont minimales),
- ☐ les redondances subjectives (il est inutile de coder les détails fins que l'œil ne voit pas).

Différents codecs sont proposés dans les normes. Ils sont différents pour l'audio et pour la vidéo.

Les codecs pour l'audio ont pour nom générique G7xx. : G711 peu utilisé, G723.1, G728, G729... Ils offrent des bandes passantes qui restent limitées aux environs de 3 KHz, ce qui ne correspond, ni plus ni moins, qu'à la qualité téléphonique. Ces codecs diffèrent par le débit nécessaire pour obtenir cette qualité (et donc par leur complexité³). G722 permet une restitution de meilleure qualité avec une bande passante atteignant 7 KHz offrant de ce fait un meilleur confort d'écoute. Différentes déclinaisons de cette norme ont été prévues. Si la première version était prévue pour des débits compris entre 48 et 64 Kb/s, la version G.722.2 permet d'obtenir la même qualité pour des débits pouvant «descendre» jusqu'à 6,6 Kb/s. Des codecs «propriétaires» peuvent également être présents, par exemple le codec TDAC développé par France Telecom qui offre une bande passante de 7 KHz.

Des nouveaux codecs plus performants ont été développés plus récemment. La société Polycom a ainsi implémenté sur ces matériels sa technologie VSX Siren 14 qui permet d'obtenir un son de qualité avec une bande passante de 14 KHz. Une restitution stéréophonique est également disponible sur certains modèles. Tandberg a pour sa part intégré un codec AAC-LD (Advanced Audio Coding Low Delay). AAC est un codage standardisé par l'ISO (International Organisation for Standardisation) et faisant partie de la spécification MPEG-4. Sa particularité est d'offrir des encodages de qualité avec des temps de calcul très courts (d'où l'appellation Low Delay) : il convient donc particulièrement bien pour des applications bidirectionnelles comme la visioconférence.

Pour la vidéo, trois normes de compression peuvent être utilisées. H261 est la plus ancienne (elle a été élaborée par le CCITT entre 1988 et 1990) et a été développée à l'origine pour une utilisation sur RNIS. La norme H263 est plus récente mais elle est aussi plus performante et

³ Pour une qualité donnée, plus le débit obtenu est faible et plus les circuits de compression doivent être performants.

a été plus particulièrement adaptée pour les communications à faibles débits. La norme H263+ est une amélioration de H263 et prend mieux en compte les spécificités des réseaux IP.

Des développements récents ont été réalisés en direction de la norme MPEG-4 et plus particulièrement vers le dernier codec vidéo, baptisé sous le double nom de H264 et Mpeg 4 part 10 (mais que l'on trouve parfois aussi sous l'appellation AVC pour Advanced Video Coding). Sur le plan qualitatif, les gains apportés sont importants. Pour des débits inférieurs à 1 Mb/s, ce codec délivre, à débit égal, un niveau de qualité d'image de 30 à 50 % supérieur à celui de la première génération de codecs MPEG-4 ASP (Advanced Simple Profile) et par comparaison de 60 % supérieur à celui du MPEG-2 qui est utilisé pour la télévision numérique ou le DVD. Appliqué à la visioconférence, il va permettre, à qualité d'image égale, une réduction importante des débits nécessaires ou en d'autres termes, pour des débits équivalents à ceux utilisés aujourd'hui une amélioration de la qualité et de la fluidité des images. On estime généralement, qu'avec une connexion de type RNIS, la qualité obtenue aujourd'hui à 256 Kb/s est identique à celle que l'on obtenait hier à 384 Kb/s avec un autre codec.

Différents fabricants ont déjà implémenté H264 sur leurs matériels. Des équipements pour micro-ordinateurs commencent aussi à intégrer cette possibilité. Le processus de normalisation de ce codec est aujourd'hui abouti (depuis juin 2004) ce qui devrait parfaire l'interopérabilité entre les différents matériels.

Pour pouvoir profiter pleinement des avantages apportés par H264, il faut bien évidemment que ce codec soit présent au deux extrémités de la chaîne, faute de quoi le système basculera automatiquement sur un autre codec, H263 par exemple. Ces innovations techniques se traduisent par une complexité accrue des dispositifs de traitement. Les ressources « machines » sont très sollicitées (on estime parfois qu'avec H264, il faut quatre fois plus de ressources machines qu'avec H263) et plus encore pour celles qui intégreront un pont multi-sites (ces équipements doivent gérer les flux issus des différents sites, les puissances de calcul doivent être multipliées d'autant). Du fait de cette complexité accrue, les équipements H264 en multipoint ne sont pas encore disponibles sur le marché.

Différents formats d'images ont été déterminés par les normes. Ils sont définis en fonction du nombre de lignes (dans le sens vertical) et de points (dans le sens horizontal) qui constituent l'image vidéo. Le format CIF (Common Interchange Format) est le format de base de la visioconférence. Il se compose de 352 points sur 288 lignes et correspond environ à un quart de l'écran sur un téléviseur.

Des formats dérivés, multiples et sous multiples, ont été également définis ainsi que le montre le tableau ci-dessous. Ces valeurs sont à comparer à celles adoptées pour une image de télévision classique (720 points x 576 lignes).

Formats		Nb de points /nb de lignes (pour la luminance)
CIF	Common Interchange Format	352 x 288
QCIF	Quarter CIF	176 x 144
SQCIF	sub quarter CIF	128 x 96
4CIF	4 x CIF	704 x 576
16CIF	16 x CIF	1408 x 1152

Pour une image de télévision, le débit traditionnel est de 25 images par seconde (30 pour le continent américain). On pourra retrouver ces valeurs sur les dépliants commerciaux fournis par les fabricants d'équipements de visioconférence (15 images par secondes au maximum sur les plus petits matériels). Pratiquement, les chiffres réels sont bien plus faibles. Il ne faut pas oublier que ces valeurs ne pourront pas être atteintes si les débits sur le réseau ne sont pas suffisants (notamment pour les réseaux IP en cas de congestion). De ce fait, une fluidité correcte des images ne sera pas toujours obtenue.

Les matériels de visioconférence disposent souvent de « sorties informatiques » qui permettent la connexion d'un moniteur informatique ou d'un vidéo projecteur pour une visualisation sur grand écran. Les résolutions graphiques habituellement proposées sont :

SXGA : 1280 x 1024

XGA : 1054 x 768

SVGA : 800 x 600

VGA : 640 x 480

Présentation et partage de documents

Une session de visioconférence ne se résume pas uniquement à des échanges oraux entre des intervenants distants, mais s'accompagne souvent de la diffusion et de la présentation de documents annexes. Proposés par l'un ou l'autre des participants, ces médias doivent pouvoir être diffusés en direction de l'ensemble des sites. Ils peuvent être de différentes natures (images fixes, photographies, séquences vidéo, fichiers informatiques, ou plus simplement, notes et schémas réalisées « en direct » sur un tableau blanc...) et proposer des degrés d'interactivité variables. Ces présentations peuvent être mises en oeuvre de différentes façons.

• La diffusion simple :

C'est la possibilité la plus simple, puisqu'il ne s'agit que de présenter un document, sans aucune interactivité possible. Pour la projection des documents vidéo, les dispositifs de visioconférence sont généralement munis d'entrées composites ou S-Video permettant la connexion d'équipements de lecture, magnétoscopes VHS par exemple. Ces entrées peuvent être également utilisées pour le raccordement d'un banc titre. Constitué d'une caméra vidéo fixée verticalement sur un piétement, cet équipement permet de présenter des petits objets ou des documents « papiers ».

Des entrées analogiques dans des formats « informatiques » courants sont également proposés pour le raccordement des micro-ordinateurs. Il ne s'agit pas ici de vouloir réaliser un travail collaboratif et interactif avec les partenaires distants mais seulement de diffuser l'image écran d'un contenu informatique via une liaison de type XGA entre l'ordinateur et le dispositif de visioconférence. C'est une solution simple qui ne nécessite qu'un seul ordinateur, et qui n'implique pas la mise en place de logiciels spécialisés.

Des dispositifs particuliers ont été développés par les industriels pour permettre la diffusion simultanée de deux sources vidéo ou d'une source vidéo et d'un flux XGA. Une nouvelle

norme H239 a été développée pour faire suite aux solutions propriétaires qui avaient été déployées (« Duo Video » pour Tandberg, « People and Content » pour Polycom). Elle est apparue au début de l'année 2004 et a été implémentée sur les premiers produits six mois plus tard. Les participants distants peuvent visualiser les deux images, soit sur un seul écran (l'une des images est en incrustation), soit sur deux écrans distincts. Pour une image informatique (diffusée via l'entrée XGA), la qualité de la restitution dépendra du type d'affichage utilisé sur le site distant. Avec un téléviseur ou un moniteur vidéo, elle pourra souffrir du transcodage (XGA vers H261 /H263) et sera affichée dans le format CIF (352 x 288) alors qu'avec un moniteur informatique elle conservera sa définition et sa qualité d'origine (matériels Aethra notamment).

Certains matériels disposent d'une fonction « PowerPoint intégrée » permettant la diffusion et la gestion d'une présentation dans ce format sans avoir recours à un micro-ordinateur extérieur. Le document est stocké en local (après transfert sous IP à partir d'un poste informatique relié au réseau, c'est le cas par exemple avec les matériels Aethra) ou sur des périphériques de stockage de type Memory Stick ou similaires (Sony). Pour en faciliter l'exploitation, l'équipement de visioconférence intègre alors toutes les fonctionnalités spécifiques nécessaires à l'exploitation d'un document de ce genre : affichage en local et en mode vignette de l'intégralité de la présentation, modification de l'ordre des diapositives, suppression... L'affichage (tout comme celui de toute image fixe, photographie ou texte) est réalisé par référence à l'annexe D de la norme H 261 qui avec une résolution de 576 lignes x 704 points permet une restitution de meilleure qualité que celle qui peut être offerte par le format CIF. Sur les sites distants, les diapositives seront présentées en plein écran (avec en incrustation l'image vidéo de l'interlocuteur) ou diffusées sur un deuxième écran.

Les dispositifs de visioconférence individuels installés sur micro-ordinateurs disposent d'un atout supplémentaire par rapport à leurs homologues de type « console » : ils intègrent par construction un élément de stockage de grande contenance (le disque dur !). Tous les documents qui y sont enregistrés peuvent être spontanément et sans doute beaucoup plus facilement diffusés vers l'ensemble des correspondants distants. C'est le cas par exemple pour le logiciel eConf développé par France Telecom R&D et de sa fonction « Drag and Stream Multimedia ». Il permet la diffusion facile de tout fichier vidéo (format mpeg, avi, mov, qt...) ou audio (wav, mp3...) avec de surcroît la possibilité d'assurer simultanément un commentaire vocal ou écrit.

Avec certains matériels (Aethra, Sony...), il sera également possible de renforcer l'impact d'une session de visioconférence par une diffusion vidéo et audio en direction du réseau IP de l'établissement ou plus généralement vers Internet. Certains équipements intègrent en effet une fonction serveur de streaming. Sous condition d'accès par mot de passe, et après connexion à l'adresse IP du serveur, elle permet à tout un chacun de suivre à distance le déroulement de la visioconférence sur son propre poste informatique. C'est bien de diffusion (audio et vidéo) dont on parle, et donc sans aucune interactivité possible, mais dans ce cadre, le nombre de « spectateurs » est illimité. Cette diffusion audio et vidéo pourra être réalisée en unicast ou en multicast (voir encadré), et, en fonction du choix opéré par le fabricant, dans l'un des formats vidéo habituellement disponible sur Internet (Real, Windows Media ou Quick Time).

Diffusion vidéo en unicast ou en multicast

Sur Internet, la diffusion en temps réel d'une séquence vidéo (ou streaming) peut être réalisée suivant deux méthodes.

La diffusion unicast correspond au concept de la vidéo à la demande : à chaque requête d'un utilisateur correspond un flux vidéo qui est délivré par le serveur. Il y aura donc sur le réseau autant de flux que de demandes. A la manière d'un magnétoscope, toutes les formes d'interactivité sont possibles.

Le multicast s'apparente plus à la diffusion télévisuelle classique telle que nous la connaissons par ailleurs (hertzienne, par satellite...). La diffusion est réalisée à un instant donné, le même pour tous, sans aucune interactivité possible. Pour visualiser la séquence, il suffit de se connecter au moment ad hoc à une adresse IP spécifique. Ce procédé permet d'optimiser la bande passante du réseau, puisque le serveur ne génère qu'un seul flux. Ce flux sera ensuite dupliqué si nécessaire au niveau de chacun des nœuds du réseau s'il se trouve en amont un récepteur valide.

● Partage de documents et fonction tableau blanc

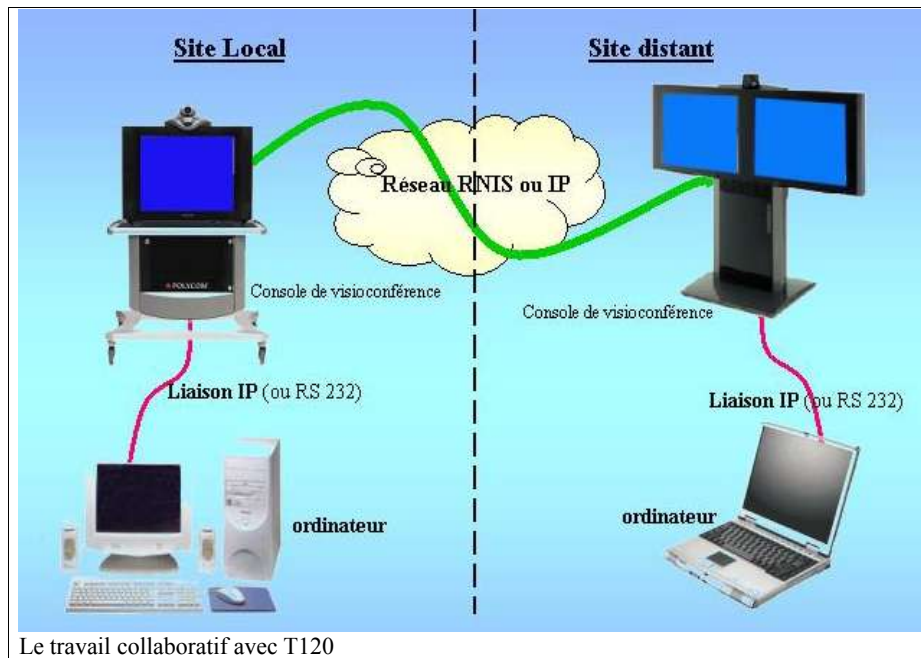
Ici, il ne s'agit plus de présenter simplement un document informatique, mais de réaliser avec les interlocuteurs distants un véritable travail interactif sur une application commune. Ce travail nécessite la connexion d'un micro-ordinateur sur chacun des sites participants.

La norme T120 est la norme de partage d'applications informatiques. Elle n'est pas spécialement dédiée à la visioconférence et fonctionne sur différents types de réseaux. C'est aussi une norme « container » car elle englobe bon nombre de protocoles spécifiques (T123, T124, T125...).

Pour réaliser le partage de documents, la grande majorité des dispositifs de visioconférence s'appuient sur le logiciel NetMeeting développé par Microsoft. C'est un logiciel déjà ancien mais qui présente l'avantage d'être disponible sur la plupart des micro-ordinateurs car intégré aux systèmes d'exploitation Windows (on évoque cependant sa probable disparition dans les versions futures). NetMeeting est un véritable outil de communication, permettant des échanges audio, vidéo ou textuels, et qui propose des fonctionnalités de type tableau blanc, partage d'application, transfert de fichiers... Seules ces trois dernières fonctions sont utilisées par les équipements de visioconférence. La fonction tableau blanc fait référence à un espace commun, visible par tous, sur lesquels les différents interlocuteurs peuvent simultanément intervenir : écrire, dessiner, coller un texte ou un tableau... Le partage d'application permet un travail collectif sur un même document (traitement de texte, tableur ou autre). Ce dernier reste hébergé sur le micro-ordinateur de son propriétaire, mais il est visible et peut être modifié à distance (si autorisation et après validation) par tous les intervenants. La fonction transfert de fichiers réalise la copie de tout document en direction des sites distants.

La mise en oeuvre d'un travail collaboratif dans le cadre d'une session de visioconférence nécessite que tous les sites soient équipés d'un micro-ordinateur sur lequel NetMeeting est actif. La liaison entre le micro-ordinateur et la console de visioconférence est établie, soit par l'intermédiaire d'une liaison IP (connexion directe ou à travers le réseau), soit via une liaison série de type RS 232 (du nom de la norme à laquelle elle fait référence). Alors même que le travail collaboratif est directement réalisé sur les micro-ordinateurs, la transmission est par contre effectuée par les équipements de visioconférence eux mêmes, en parallèle avec

les signaux audio, vidéo, de contrôle... propres à la session. Cette transmission s'appuie sur le protocole TCP, car il est nécessaire de conserver l'intégrité des documents transmis (les paquets perdus doivent être ré-émis).



- Le couplage avec un tableau blanc numérique :

Dispositif indépendant des systèmes de visioconférence, un tableau blanc interactif est un équipement spécifique de visualisation destiné à une utilisation collective. Interconnecté à un micro-ordinateur et à un vidéo projecteur, il permet non seulement la projection du contenu de l'écran informatique (quel qu'en soit le contenu), mais offre également (en sur-impression sur le document projeté ou sur une surface blanche mais dans les deux cas à la manière d'un tableau classique) des fonctions d'écriture, de dessin, d'annotation, de surlignage... La spécificité de cet équipement est de permettre la sauvegarde de tous les travaux qui y sont réalisés et d'assurer le pilotage à distance du micro-ordinateur.

Ces dispositifs ont été conçus pour une utilisation en local dans une salle de formation, mais ils peuvent être également couplés à des équipements de visioconférence. L'ensemble des opérations réalisées sur la tableau sont alors totalement visualisables par l'ensemble des interlocuteurs distants.

- Web conférences

Une Web conférence est une réunion virtuelle établie via Internet (ou via un réseau IP) entre différents interlocuteurs situés sur des zones géographiques distantes. Ces dispositifs ne sont conçus que pour des utilisations individuelles à partir d'un micro-ordinateur et offrent généralement des fonctionnalités de messagerie, de communications interactives instantanées (chat...), de présentation de documents, de navigation sur le Web, de tableau blanc... Les documents à présenter doivent avoir été placés sur un serveur spécifique qui pourra être éventuellement hébergé par une société tierce, et auquel tous les participants pourront facilement se connecter (par exemple, à l'aide d'un simple navigateur). L'initiateur

de la session conserve la gestion de l'organisation (invitation, définition des mots de passe) et la maîtrise du déroulement de la réunion (diffusion d'un document, suivi des interventions...). Différentes fonctionnalités supplémentaires peuvent être proposées : prise de contrôle à distance, enregistrement de la session...

Ces dispositifs ont été conçus indépendamment de tout système de visioconférence et n'incluent d'ailleurs pas nécessairement des outils vidéo ou audio. Ils peuvent cependant intégrer une liaison audio entre les participants, soit directement via les outils disponibles sur le poste informatique, soit par l'intermédiaire d'une communication téléphonique classique établie par ailleurs. D'autres peuvent également offrir la retransmission de l'image vidéo de la personne qui a la parole.

Ces dispositifs peuvent venir en complément à des solutions complètes de visioconférences qui apporteront alors des liaisons audios et vidéo de bien meilleure qualité.

Nouvelles normes et évolutions

- Centraliser les adresses avec la nouvelle norme H350

Cette nouvelle norme (elle date de septembre 2003) concerne la fonction d'annuaire et normalise les procédures de stockage des données d'adresses (qui étaient souvent propriétaires). Elle permet la centralisation sur un serveur spécifique de toutes ces informations, non seulement les adresses IP ou les alias des terminaux H323 de visioconférence, mais également celles utilisées par d'autres types d'équipements, notamment ceux destinés à la téléphonie sur IP. La recommandation H350 s'appuie sur le protocole LDAP (Lightweight Directory Access Protocol). Elle est compatible H320, H323, SIP...

- SIP (Session Initiation Protocol)

SIP est à l'origine un protocole téléphonique sur lequel on a ajouté aujourd'hui la possibilité de transmettre de la vidéo. Pour les équipements de visioconférence, c'est l'une des dernières tendances pour cette fin d'année 2004. La norme SIP a été implémentée sur des produits Polycom depuis le mois de juillet 2004.

SIP est plus récent que H323. Cette norme a été développée et normalisée sous la tutelle de l'ITU-T (le monde des télécoms !) par opposition à H323 qui a été développé sous l'égide de IETF (le monde de l'Internet !). Elle s'implante largement dans les entreprises grâce au succès croissant de la téléphonie sous IP (ou VoIP). Pour la visioconférence, ce protocole n'a cependant pas la richesse de H323 et il offre des possibilités en terme de fonctionnalités plus réduites (il ne permet pas de faire du multipoint directement, il présente des insuffisances pour la commande à distance des caméras...). Mais il présente l'avantage de pouvoir s'appuyer sur les réseaux téléphoniques d'entreprise sur IP pour la mise en œuvre des dispositifs de visioconférences. SIP peut donc constituer une alternative d'avenir à H323.

- WiFi (Wireless Fidelity) :

Il ne s'agit pas ici d'une fonctionnalité spécifique aux équipements de visioconférence mais simplement de la mise en œuvre d'une liaison sans fil en lieu et place de la liaison filaire classique nécessaire entre le terminal de visioconférence et le réseau IP. WiFi est une norme de transmission radio (802.11) qui permet l'échange de données entre deux périphériques. Plusieurs déclinaisons ont été réalisées. La plus connue est la norme 802.11b qui utilise la

plage de fréquence de 2,4 GHz et permet un débit théorique de 11 Mb/s. Plus récente, la norme 802.11g autorise des débits qui peuvent atteindre 54 Mb/s. Dans les deux cas, les liaisons sont bi-latérales mais restent limitées à de courtes distances (100 m en champ libre mais beaucoup moins à l'intérieur d'un bâtiment).

L'équipement de visioconférence est équipé d'un module spécifique d'émission et réception. Une borne d'émission réception reliée au réseau local de l'établissement doit être présente à faible distance.

- La confidentialité avec le cryptage AES (Advanced Encryption Standard) :

Les matériels de visioconférence offrent désormais des dispositifs d'encryptage qui permettent d'assurer la confidentialité des échanges, avec des communications qui sont sécurisées et chiffrées. C'est une possibilité qui est proposée pour les entreprises « sensibles » mais qui ne concerne pas directement le monde de l'éducation. Cette fonctionnalité est totalement transparente pour l'utilisateur. Différentes normes existent H233, H234... la plus récente étant H 235.

Sous IP, des contraintes liées aux réseaux locaux et à leurs équipements spécifiques

Parallèlement aux difficultés inhérentes à Internet et déjà évoquées dans les chapitres précédents, la visioconférence sous IP souffre également des maux liés à certains dispositifs informatiques spécifiques mis en place dans les réseaux locaux d'entreprise. Situés à la frontière entre le réseau local et le réseau extérieur, ces dispositifs sont principalement de deux ordres : ceux destinés à assurer la gestion des adresses IP de l'ensemble des micro-ordinateurs connectés au réseau local et ceux destinés à assurer la sécurité et la protection des connexions (FireWall). Dans le cadre d'une visioconférence, ces équipements induisent des difficultés d'exploitation particulières qu'il est nécessaire de bien connaître.

La gestion des adresses

Sur Internet tout comme sur les réseaux locaux de type Ethernet, chaque poste informatique se voit affecter un identifiant unique appelé adresse IP. Deux postes sur un même réseau (Internet étant considéré comme l'un d'entre eux) ne peuvent pas avoir la même adresse IP. Ces adresses sont codées sur 4 octets de 8 bits (soit 32 bits)⁴ dans la version actuelle du protocole IP et leur nombre est par conséquent limité. Cette limitation devrait disparaître avec la future version IPv6.

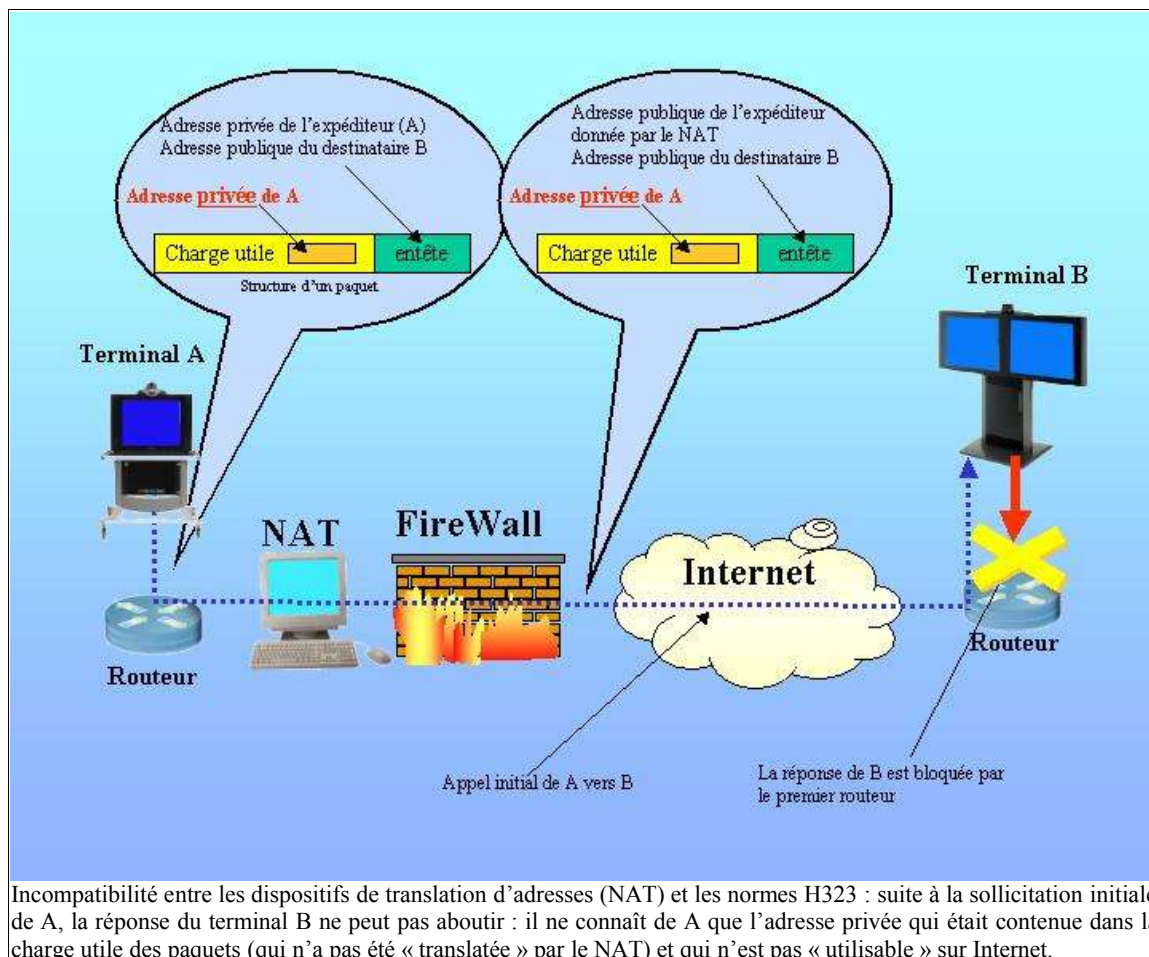
La technique de translation d'adresse (NAT en anglais pour Network Address Translation) a été mise au point pour gérer la pénurie d'adresses possibles face aux besoins croissants d'Internet. Il a ainsi été décidé de répartir le volume des adresses possibles en deux sous ensembles : les adresses publiques (utilisables sur Internet) et les adresses privées (employables uniquement à l'intérieur des réseaux privés). Sur les réseaux locaux, ces dernières peuvent être utilisées sans d'autre restriction que ne pas attribuer deux fois la même adresse dans un même réseau local. Compte tenu de ce confinement, il n'y a aucun risque de conflit lorsqu'une même adresse privée est utilisée sur des réseaux locaux différents.

Sachant qu'une adresse privée ne peut pas être utilisée sur Internet, lorsqu'un poste souhaite se connecter, le mécanisme de translation d'adresse va remplacer l'adresse privée présente dans l'entête de chacun des paquets par une adresse publique avant de router ensuite le paquet vers l'extérieur. Il réalisera l'opération inverse au retour de la réponse. Cette translation pourra être statique (à chaque adresse privée correspondra toujours la même adresse publique) ou dynamique (il n'y a pas d'associations prédéfinies). Dans ce cas, l'établissement disposera en général de moins d'adresses publiques au regard du nombre de postes réellement présent dans le réseau local. Elles seront attribuées par le système au fur et à mesure des demandes de connexion. Plusieurs utilisateurs pourront se voir attribuer la même adresse, la différenciation au retour entre les données destinées aux uns et des autres et la détermination du poste émetteur original s'effectuant alors sur d'autres critères.

Outre le fait que la technologie de la translation dynamique d'adresses permet de limiter le nombre d'adresses publiques utilisées par un réseau s'ouvrant sur l'extérieur, elle permet aussi d'assurer la protection des machines internes contre des « actions malveillantes en provenance d'Internet » puisque leur adresse IP réelle n'est en fait pas directement « visible de l'extérieur ». Cette fonction de sécurité est la deuxième raison de l'existence des NAT.

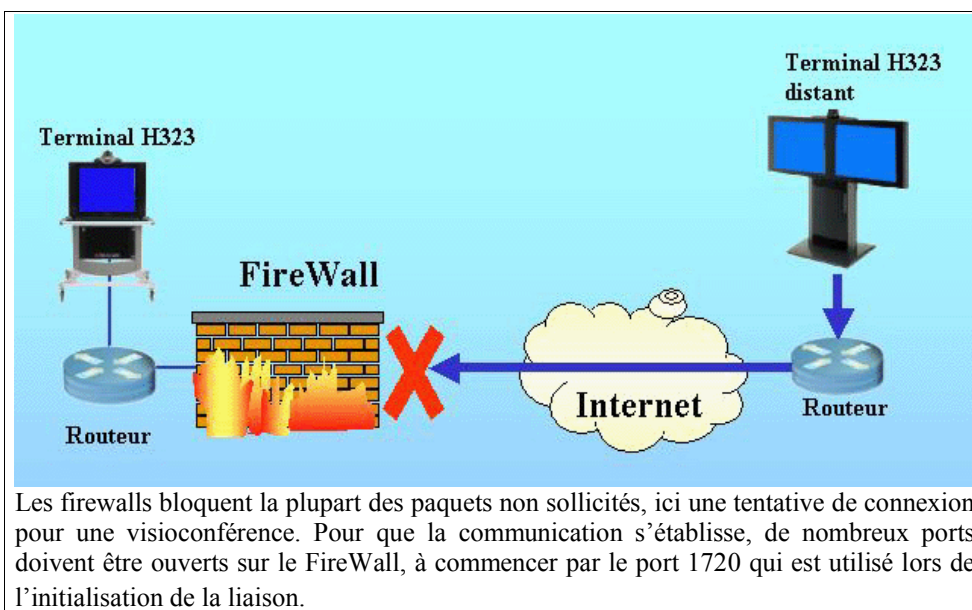
⁴ Elles sont indiquées par quatre nombre compris entre 0 et 255 séparés par un point (par exemple 194.250.164.58 est l'une d'entre elles) ce qui représente environ 4,3 milliards d'adresses possibles.

Ce dispositif n'est malheureusement pas pris en compte par certaines des normes (H225 ou H245) utilisées dans le cadre des visioconférences. Ces deux normes n'exploitent pas les adresses contenues dans les entêtes, mais elles inscrivent (et utilisent ensuite) des informations d'adresse qu'elles placent dans le corps même du paquet (dans le schéma ci-dessous, c'est l'adresse privée du terminal A qui va être ainsi transmise au terminal B lors de l'appel initial de A). Compte tenu de sa localisation dans la charge utile, cette adresse n'est pas transformée par le NAT. En tentant d'exploiter cette information pour répondre, l'équipement de visioconférence distant ne peut trouver qu'une adresse privée totalement inexploitable sur Internet. L'établissement de la session de visioconférence est impossible. Certains équipements de visioconférence intègrent des fonctionnalités spécifiques (fonction « NAT IP Adress mask » pour Vcon, « Aethra NAT » pour Aethra...) qui permettent de pallier à cet état de fait (via un menu de configuration, l'utilisateur peut indiquer manuellement l'adresse IP publique qui sera systématiquement attribuée). L'utilisation de firewall ou de proxy compatibles H323 (voir chapitre suivant) sont aussi des méthodes pour résoudre ces cas de figure.



Le passage des firewalls

Un Firewall (pare feu dans la littérature française) est un dispositif de sécurité placé à la jonction entre deux réseaux distincts, le réseau informatique interne à un établissement et le réseau extérieur, Internet en l'occurrence. Organe de sécurité destiné à protéger le réseau interne, sa tâche principale est d'interdire les activités malveillantes en provenance de l'extérieur. La contrainte du firewall est d'être le plus transparent possible pour les activités à l'intérieur de l'entreprise et d'être à la fois le plus efficace possible en offrant un niveau maximum de sécurité. C'est essentiellement un outil de filtrage destiné au contrôle de la circulation des paquets, et qui doit assurer le blocage de toutes les données qui ne doivent pas passer d'un côté à l'autre. Il va généralement interdire toutes les «entrées» de données qui ne répondraient pas à une requête préalable de l'un des postes du réseau local. Comment dès lors, répondre à une demande d'initialisation pour une session de visioconférence lorsqu'elle est sollicitée depuis l'extérieur ?



Un firewall va également assurer la surveillance des « ports » qui sont utilisés. Sur un micro-ordinateur, chaque application logicielle se voit attribuer un port (le port est en quelque sorte « l'adresse » d'une application). Lors d'une connexion « classique » à Internet la majorité des ports sont fermés sur le firewall, seuls les quelques uns qui correspondent aux applications directement concernés sont ouverts. Dans le cadre de la visioconférence, de nombreuses connexions doivent être simultanément maintenues entre les terminaux, et de nombreux ports doivent y être ouverts, certains aléatoirement (c'est à dire sans que l'on puisse prévoir préalablement leur numéro). Cette notion de ports dynamiques ne facilite pas la configuration des firewall : pas question de laisser tous les ports entre 1024 et 65535 ouverts ! Sauf mise en place de dispositifs particuliers, l'ouverture de tous ces ports sont autant de failles dans la sécurité globale d'un réseau local. A l'inverse, du fait des dispositifs de protection adoptés par les administrateurs de réseau, la mise en place de séances de visioconférence peut se révéler difficile, parfois même impossible.

Pour en savoir plus

Sur un micro-ordinateur, chaque application logicielle se voit attribuer un port (le port est en quelque sorte « l'adresse » de l'application). Pour des données en provenance de l'extérieur, le numéro de port indique à quelle application sont destinées les données. Les ports sont codés sur 16 bits, 65535 ports sont théoriquement disponibles, pratiquement moins, car 1024 sont réservés.

Lors d'une connexion « classique » sur Internet la majorité des ports sont fermés, seuls les quelques uns qui correspondent aux applications directement concernés sont ouverts pour permettre les échanges de données (port 80 pour HTTP, ports 25 et 110, respectivement pour les échanges SMTP et POP3 de la messagerie...). Bloquer l'utilisation d'un port, c'est interdire le transit des données correspondant à certaines applications.

Dans le cadre de la visioconférence, certains de ces ports sont spécifiés d'une manière définitive (ports statiques) par la norme H263, par exemple, port 1720 pour l'appel initial, port 1719 en cas d'utilisation d'un gatekeeper, port 1503 pour le partage d'applications via la norme T120... D'autres (ports dynamiques) sont attribués aléatoirement au moment de l'établissement de l'appel (ports compris entre 1024 et 65535). Ce sont par exemple ceux utilisés pour le transfert des données vidéo et audio (flux RTP et RTCP).

Suivant le type de données, les transferts pourront s'effectuer en utilisant les protocoles TCP ou UDP.

Quelques exemples des ports utilisés pour une visioconférence :

Port	Type	Protocoles	Description
1719	Statique	UDP	Gatekeeper RAS
1720	Statique	TCP	Q.931 (Call Setup)
1024-65535	Dynamique	TCP	H245 (Call parameters)
1024-65535	Dynamique	UDP (RTP)	Video and audio Data Stream
1024-65535	Dynamique	UDP (RTCP)	Control Video and audio Stream
Ports optionnels			
389	Statique	TCP	ILS Registration (LDAP)
1503	Statique	TCP	T.120

Différentes solutions techniques ont été développées pour contourner ces obstacles et permettre un fonctionnement correct des protocoles H323 à travers les firewalls :

- Utiliser des firewalls intégrant H323 :

C'est sans doute la meilleure solution, et celle qui offre la meilleure sécurité. Beaucoup des firewalls récents intègrent désormais H323 (sous l'appellation Application Level Gateways ou ALG dans certains textes). Ces équipements ont la faculté de scruter les communications qui sont établies en amorce à une visioconférence afin de détecter les numéros de ports qui seront effectivement utilisés. Ils pourront alors autoriser l'ouverture de ces ports spécifiques et permettre le trafic entre appelé et appelant pendant une durée qui restera limitée à celle de la session. Ces ports sont refermés ensuite. On utilise parfois le terme de « pinholing » pour désigner cette méthode qui consiste à n'ouvrir que les quelques ports nécessaires (des « trous d'épingle ») dans le firewall.

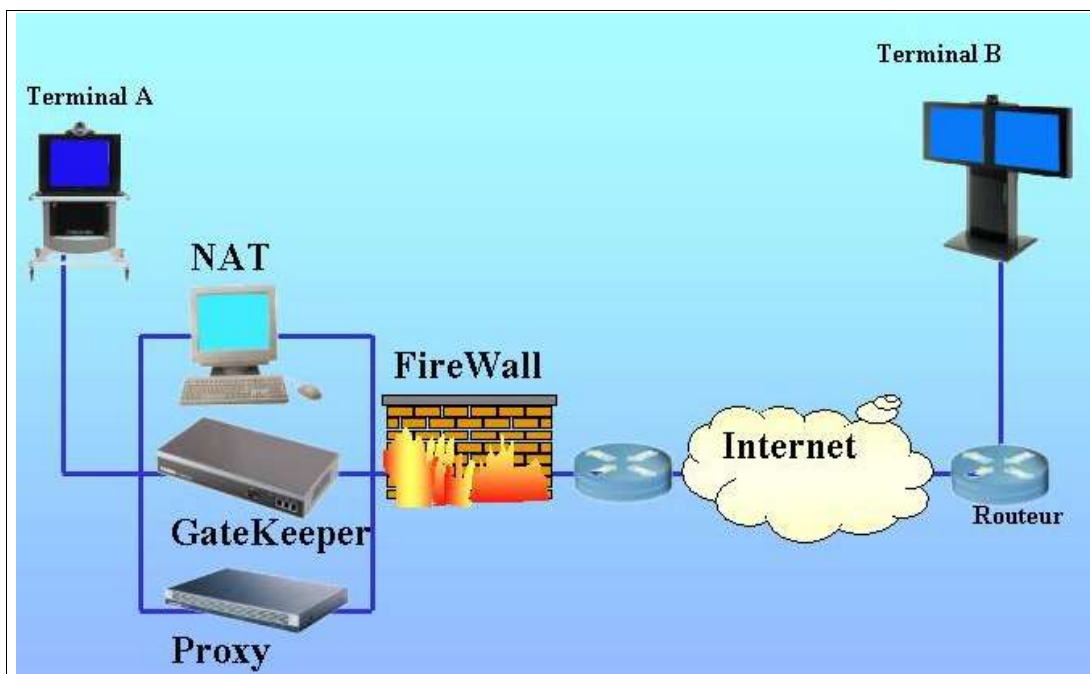
Certains modèles intègrent également la fonction NAT. Lors de la translation d'adresses, ils sont capables, non seulement de remplacer une adresse privée par une adresse publique dans l'entête des paquets, mais également de réaliser cette opération dans le corps même de la charge utile, permettant de ce fait le fonctionnement correct de toute session de visioconférence.

- Utiliser des proxys

Un proxy est une passerelle spécialisée qui va permettre à des flux H323 de contourner dans certaines conditions les firewalls, sans affaiblir les conditions de sécurité. Il va agir comme un intermédiaire. C'est lui qui va assurer la gestion de tous le trafic H323 en lieu et place des terminaux de visioconférences qui seront ainsi totalement isolés d'Internet (ils seront invisibles de l'extérieur, y compris pont et passerelle).

Lors de l'établissement d'une liaison, ce n'est plus un appel qui sera généré mais deux. Le premier sera initié par l'équipement de visioconférence situé à l'intérieur du réseau local en direction du proxy qui à son tour en générera un second sur le réseau public (et en utilisant sa propre adresse) en direction de l'équipement distant. Seul le proxy peut inter-agir avec l'extérieur. Le Firewall devra être correctement configuré pour pouvoir fonctionner lui. Ce mode de fonctionnement impose l'utilisation d'un gatekeeper (voir chapitre suivant).

Différentes configurations sont possibles pour le proxy : il pourra être intégré au gatekeeper ou au firewall.



Un proxy est une passerelle spécialisée qui va permettre à des flux H323 de traverser dans certaines conditions les firewall

- S'appuyer sur une « zone démilitarisée » ou DMZ :

Une DMZ est une zone particulière du réseau informatique d'un établissement ou d'une entreprise. C'est une zone séparée qui ne va héberger que les équipements qui doivent être accessibles depuis l'extérieur, non seulement les serveurs (serveurs Web, serveurs FTP, serveurs Email...) mais également éventuellement des dispositifs de visioconférence. Elle est située, sur le plan des risques, entre le réseau local privé (qui doit être totalement protégé) et Internet (qui est une zone à très fort risque). La DMZ doit être également accessible depuis le réseau privé. Les adresses pourront y être privées ou ce qui est beaucoup mieux, publiques pour permettre des accès sans translation d'adresse. Les règles de communications entre les trois entités - le réseau local, la DMZ et Internet - seront différentes et gérées par un Firewall.

Un équipement particulier : le Gatekeeper

La fonctionnalité de NAT n'a pas que des avantages. Pour la visioconférence, elle ne permet pas à un utilisateur distant et extérieur de se connecter facilement (il ne connaît pas l'adresse IP de son interlocuteur). Il est nécessaire d'ajouter au dispositif un élément supplémentaire, le gatekeeper (garde barrière, en abrégé GK dans la littérature concernée) qui sera chargé en dépit des procédures appliquées aux adresses IP d'assurer la communication entre les différents sites et la mise en liaison des différents intervenants.

Un GK est un serveur spécifique qui va tenir à la fois un rôle de central téléphonique, d'annuaire et de gestionnaire. Il assure la gestion de tous les terminaux situés dans une zone déterminée, les met en liaison les uns avec les autres, dans la même zone ou vers d'autres zones.

Les GK sont optionnels (deux terminaux peuvent très bien communiquer entre eux directement) mais ils sont nécessaires si on souhaite développer des fonctionnalités plus complètes dans une zone.

Le gatekeeper permet d'identifier et de référencer les intervenants sous une forme plus explicite que l'adresse IP. Il est possible d'utiliser un nom et un prénom, une adresse Email ou toute autre indication. On utilise le terme d'alias pour désigner ce référencement alphanumérique. Il est aussi possible d'utiliser un indicatif numérique sensiblement identique à celui d'un numéro de téléphone (E164). Le GK établit et mémorise la relation entre l'adresse IP de la machine et la dénomination sous laquelle elle a été référencée. Il s'agit ici de traduction d'adresse et non pas d'une translation comme pour le NAT. Tout comme une adresse IP, un alias doit être unique. Lorsque un poste veut s'intégrer dans une visioconférence et joindre un correspondant, il lui suffit d'indiquer l'alias du correspondant. L'appel transitera par le GK situé dans la zone. Si le correspondant recherché est localisé dans une autre zone, il transmettra la demande vers d'autres GK.

Encadré :

Sur un réseau local (mais c'est aussi souvent le cas pour le particulier qui se connecte à Internet de chez lui à travers son FAI), l'adresse IP d'un poste n'est pas toujours fixe et attribuée une fois pour toute (adresse statique). Dans certains cas, chaque ordinateur se voit attribuer une adresse provisoire au moment de son allumage. C'est le protocole DHCP (Dynamic Host Configuration Protocol) qui assure automatiquement (et d'une manière totalement transparente pour l'utilisateur) cette fonction d'attribution et d'administration des adresses dans le réseau local. Dans ce cas, le GK doit être mis à jour en permanence : à chacune des ouvertures du logiciel de visioconférence (pour un ordinateur individuel), il y a transmission à l'annuaire de l'adresse IP de la machine concernée (elle s'enregistre auprès du GK).

Après l'établissement de la liaison et la mise en oeuvre des procédures préliminaires entre les différents sites, le GK peut, soit être mis hors du circuit de la visioconférence (les données vont alors transiter directement entre les interlocuteurs) soit, au contraire, assurer aussi le transit partiel (uniquement les données de contrôle) ou total de toutes les communications (routage complet).

Le rôle du GK ne se limite pas à la fonction d'annuaire. Il peut également réaliser la gestion des flux ou plus précisément de la bande passante affectée à une visioconférence. Il pourra par exemple autoriser ou non une connexion en fonction de la charge du réseau si cela risque d'engorger le réseau interne, limiter le débit pour un utilisateur ou pour un groupe d'utilisateurs, limiter le nombre de terminaux H263 simultanément en fonction sur le réseau, gérer le contrôle d'accès et refuser des connexions....

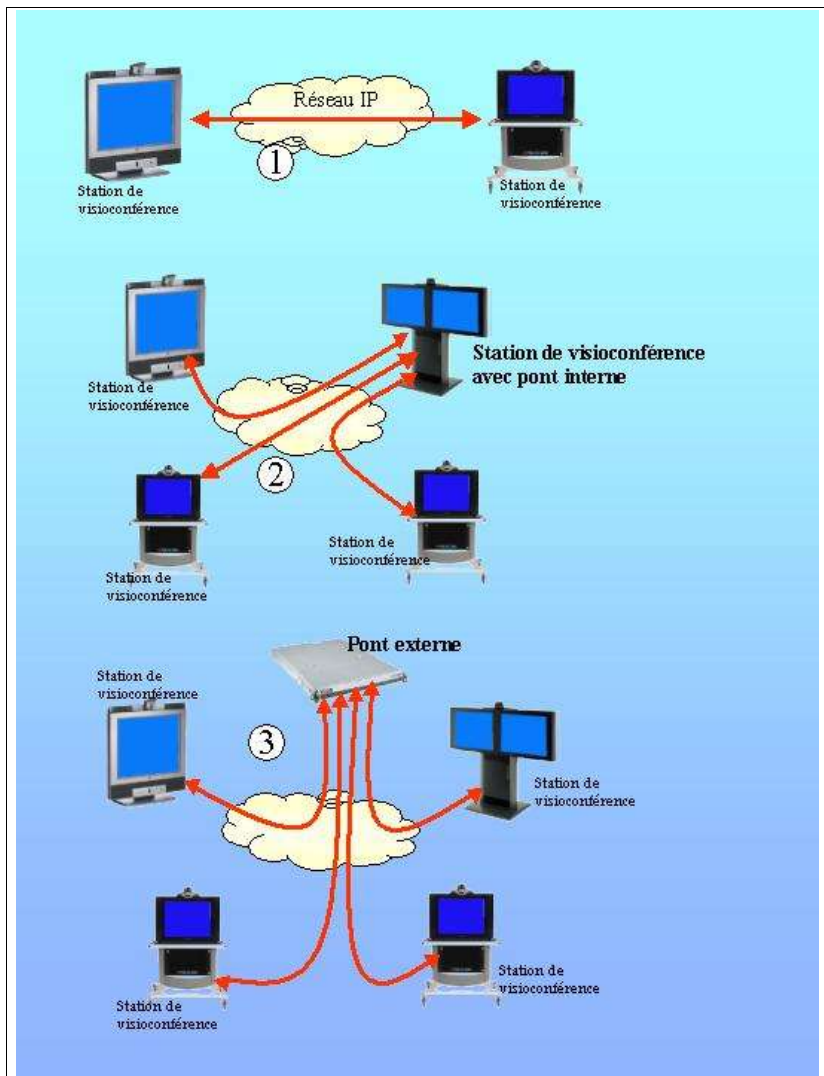
La visioconférence en multi point

Une séance de visioconférence se déroule rarement entre seulement deux intervenants. Plusieurs sites sont en général concernés. Des dispositifs spécifiques doivent être mis en place pour permettre à tous les intervenants de travailler dans les mêmes conditions : chaque site doit recevoir les images et le son en provenance de chacun des autres lieux. Il doit pouvoir visualiser un seul site à la fois (celui qui prend la parole est affiché en plein écran) ou plusieurs sites en simultané (l'écran est divisé), intervenir dans la conversation, travailler en temps réel sur des documents informatiques...

Hors réseaux universitaires (ils disposent d'autres possibilités), on utilise généralement un équipement supplémentaire auquel tous les sites doivent être connectés (pont multi-points ou MCU Multipoint Conferencing Unit). Cet équipement centralise les flux issus de tous les sites et assure ensuite leur re-distribution vers l'ensemble des équipements participant à la visioconférence. Ce sont donc des liaisons point à point qui sont établies entre le pont et chacun des postes participants. Au moment de l'initialisation de la session de visioconférence, soit c'est le pont qui appelle les participants, soit c'est l'inverse.

Un pont peut être autonome ou intégré dans l'une des stations. Dans le premier cas, c'est le pont (et le réseau auquel il est relié) qui supportera la multiplication des flux, chacune des stations ne supportant qu'un flux unitaire et somme toute équivalent à ce qu'il serait pour une liaison classique en point à point (par exemple 256 Kb/s). Dans le second cas, tous les flux convergeant vers la station qui intègre le pont, elle aura (ainsi que le réseau local sur lequel elle est située) à gérer un flux qui sera multiplié par le nombre d'interlocuteurs extérieurs (dans notre exemple, 256 Kb/s multiplié par 3 soit 768 Kb/s pour une visioconférence entre 4 points)

Un pont externe peut être acheté ou loué séparément à la demande à une société tierce. Il peut être mixte RNIS et IP. Un pont MCU peut gérer simultanément plusieurs visioconférences (sauf pour les modèles bas de gamme).



Les trois configurations possibles pour une visioconférence :
 en 1, liaison point à point
 en 2, visioconférence entre quatre points avec un pont intégré
 en 3, session entre quatre points avec un pont externe.

Interconnecter les deux familles RNIS et IP : les passerelles

Le monde de la visioconférence est partagé entre les équipements qui fonctionnent sur RNIS et ceux qui s'appuient sur IP. Il était nécessaire de pouvoir établir des liaisons entre ces deux entités.

Une passerelle (gateway dans la littérature anglaise) est un équipement permettant la communication entre une zone de visioconférence fonctionnant sous RNIS et un site fonctionnant sous IP, c'est à dire entre des terminaux qui répondent à la norme H323 et des terminaux qui s'appuient sur H320.

Entre ces deux familles de normes, bon nombre de protocoles sont différents. La passerelle va assurer leur translation, notamment pour ceux qui concernent les formats de transmission

(c'est à dire H225 en H221) ou les procédures de communications (H245 en H242). La passerelle procédera aussi à la transformation des adresses des terminaux (adresses sous la forme IP pour la norme H323, adresses dans un format « téléphonique » pour la norme H320) mais elle ne réalisera généralement pas de translation au niveau des données vidéo et audio, les codecs utilisés étant communs aux deux systèmes.

La passerelle va également assurer une remise en forme de toutes les données. Sous RNIS, elles circulent en flux continu alors que sur IP, elles sont découpées en paquets. Lorsque les données vont transiter du monde RNIS au monde IP, la passerelle va réaliser le découpage du flux continu de données et procéder à la création des paquets. Elle va y ajouter des entêtes (ces données supplémentaires qui n'existent pas sous RNIS, qui contiennent notamment les informations nécessaires pour l'acheminement des paquets et qu'il faut également transmettre). Ces données supplémentaires vont générer un débit supplémentaire qui pourra atteindre 20 à 30% du débit initial. A qualité égale une visioconférence sur réseau IP consommera donc une bande passante supérieure à celle utilisant RNIS. Globalement, on admet généralement qu'il faut un débit de 384 Kb/s sous IP pour obtenir une liaison équivalente à une visioconférence avec RNIS à 256 Kb/s. La passerelle effectuera évidemment l'opération inverse pour les données circulant dans le sens IP vers RNIS