La diffusion multipoint (le multicast), le MBone

Christian Claveleira, claveleira@univ-rennes1.fr

Octobre 1995

Résumé

Le multimédia et le travail collaboratif sont en vogue actuellement mais ils ne connaitraient pas cet essor sans le support des techniques de diffusion multipoint... Cet article présente succintement les concepts des transmissions en mode multipoint, les évolutions d'adressage et de protocoles de routage qui permettent leur utilisation et le réseau support de ces transmissions, le MBone, avec la façon de s'y connecter.

1 Introduction

La notion de transmission de données en mode point à point (terminal-ordinateur, client-serveur) a évoluée pour faire face à de nouveaux besoins : la communication de un vers plusieurs ou plusieurs vers plusieurs.

Cette notion de transmission multipoint (multicast) peut être implémentée au niveau applicatif, comme cela est réalisé, par exemple, lorsque l'on envoie un courrier électronique à une liste de destinataires ou que l'on poste une intervention dans un groupe de "news". Mais cela devient rapidement lourd à gérer dès que la liste de destinataires varie. Cette approche ne permet que très difficilement des applications de diffusion de son et/ou de vidéo auxquelles chacun pourrait se "connecter" librement. D'autre part cela peut conduire à faire circuler de multiples exemplaires des mêmes données sur un même lien, consommant ainsi de la bande passante. Des extensions ont donc été apportées au niveau de la couche IP, de nouveaux protocoles de routage ont été développés et un réseau expérimental de diffusion multipoint se déploie : le MBone.

2 L'implémentation du multipoint IP

Aux trois classes d'adressage IP traditionnelles (A, B et C) s'ajoute la classe D dite multipoint. Toute adresse IP commençant par "1110" appartient à cette classe et représente une adresse de groupe. Un tel groupe représente un nombre quelconque (y compris nul) d'équipements sur le réseau. En notation habituelle les adresses de groupe vont de 224.0.0.0 à 239.255.255.255. Deux adresses sont réservées : 224.0.0.0 qui ne doit pas être utilisée et 224.0.0.1 représentant l'ensemble des machines IP sur le réseau local. Les groupes évoluant dynamiquement, les systèmes rejoignent et quittent à la demande un nombre quelconque de groupes. A noter qu'il n'est pas nécessaire d'être membre d'un groupe pour émettre des données vers ce groupe.

2.1 Le protocole de gestion de groupe IGMP

IGMP (Internet Group Management Protocol), défini dans le RFC 1112, permet aux machines de déclarer leur appartenance à un ou plusieurs groupes auprès du routeur multipoint dont elles dépendent soit spontanément soit après interrogation du routeur. Celui-ci diffusera alors les datagrammes destinés à ce ou ces groupes. IGMP, comme ICMP, fait partie de IP (protocole numéro 2) et comprend essentiellement deux types de messages : un message d'interrogation (Host Membership Query), utilisé par les routeurs, pour découvrir et/ou suivre l'existence de membres d'un groupe et un message de réponse (Host Membership Report), délivré en réponse au premier, par au moins un membre du groupe concerné.

Les implémentations actuelles d'IGMP apportent un certain nombre d'améliorations par rapport au RFC, en particulier sur la rapidité de détection d'absence d'abonné à un groupe sur un réseau.

2.2 Multipoint sur réseau Ethernet

L'adressage multipoint étant défini au niveau de la couche réseau, il reste à la couche MAC à délivrer les trames. Dans le cas d'Ethernet la notion d'adressage de groupe existe depuis sa spécification dans l'avis IEEE802.3. Pour transporter des datagrammes IP multipoint on construit l'adresse Ethernet de destination de la façon suivante : les 23 bits de poids faible de l'adresse de groupe IP sont mis dans les 23 bits de poids faible de l'adresse Ethernet 01-00-5E-00-00-00. Cette correspondance n'est pas univoque puisqu'il y a 28 bits significatifs dans l'adresse IP mais cela ne pose pas de problème dans la pratique.

3 Le routage multipoint

Des protocoles de routage spécifiques ont été développés ou sont en cours de spécification pour résoudre les problèmes suivants :

- comment atteindre les membres des différents groupes répartis sur tout l'Internet (construction d'arbres d'acheminement)
- comment économiser de la bande passante en n'acheminant les paquets multipoint que là où il y a des membres des groupes correspondants
- comment optimiser les échanges entre routeurs (vaut-il mieux annoncer quels sont les groupes que l'on souhaite recevoir ou ceux que l'on ne veut pas recevoir ?)

On peut actuellement répartir les protocoles de routage multipoint en deux familles :

- ceux orientés "forte densité de clients" (dense-mode) comme DVMRP (Distance Vector Multicast Protocol) qui est au multipoint ce que RIP est au monopoint (unicast), MOSPF qui est l'extension multipoint de OSPF et PIM-DM, le plus récent. Ces protocoles supposent qu'il y a des membres des groupes multipoint sur la plupart des réseaux et que l'absence de membre constitue l'exception pour laquelle il y aura transfert d'information entre routeurs.

– ceux orientés ”faible densité de clients” (sparse-mode) comme CBT (Core Based Tree) et PIM-SM. Ces protocoles supposent, au contraire des précédents, que les membres de groupe multipoint sont très dispersés et peu nombreux par rapport au nombre de réseaux desservis.

3.1 DVMRP

Décrit dans le RFC1075, DVMRP est dérivé du protocole RIP (Routing Information Protocol) et, comme lui, utilise la notion de distance. Les éléments de protocole sont mis dans des paquets IGMP. L'algorithme utilisé dans DVMRP est le Truncated Reverse Path Broadcasting (TRB) dont l'objet est de déterminer le chemin le plus court entre la source et tous les receveurs. Chaque routeur détermine sa place sur l'arbre ainsi constitué afin de déterminer lequel de ses interfaces se trouve sur le chemin le plus court. Les feuilles de l'arbre peuvent ne plus avoir de receveurs pour un groupe donné, l'arbre est alors tronqué (prunning).

DVMRP inclut la notion de tunneling permettant de propager des paquets multipoint à travers des routeurs ne supportant pas le routage multipoint. Pour cela les paquets multipoint ont un en-tête IP un peu particulier : les adresses sources et destination originelles sont mises dans deux éléments de ”loose source routing” (routage à la source lâche), l'adresse source est celle du routeur qui émet le paquet et l'adresse destination est celle du routeur multipoint destinataire (à l'autre bout du tunnel). Celui-ci rétablit alors les adresses sources et destination d'origine.

DVMRP a été implémenté sous la forme du démon mrouted pour système Unix, à la base du MBone actuel (cf ”4.0 mrouted”) ainsi que sur certains routeurs (Proteon). Mais, comme tous les protocoles basés sur la notion de distance, il se prête mal à un déploiement à grande échelle.

3.2 PIM (Protocol Independent Multicast)

Issu du groupe IDMR (Inter-Domain Multicast Routing) de l'IETF, PIM se veut plus efficace que DVMRP ou MOSPF pour un large déploiement : pour ces derniers, lorsque les membres d'un groupe et les émetteurs vers ce groupe sont clairsemés sur une grande région, des paquets de données sont envoyés régulièrement sur de nombreux liens qui n'aboutissent ni à des membres ni à des émetteurs. Le mode SP (Sparse Mode) de PIM permet d'y remédier en n'envoyant des messages qu'en cas de rattachement à un groupe (et non le contraire) et en utilisant un mécanisme de rendez-vous entre émetteurs et membres de groupes. Le mode DM (Dense Mode) apporte également des améliorations par rapport à DVMRP ou MOSPF.

Les deux modes de fonctionnement (dense et sparse) coexistent avec adaptation dynamique groupe par groupe.

PIM est décrit dans les documents de travail de l'IETF (deuxième version) : draft-ietf-idmr-pim-*.ps et est implémenté actuellement sur les routeurs Cisco.

4 mrouted

Le démon mrouted est initialement une implémentation de DVMRP (RFC1075) sous Unix et constitue le coeur actuel du MBone (cf ”5.0 Le MBone”). Cette implémentation a évoluée au point qu'il y a aujourd'hui de nombreuses différences avec le RFC. En particulier l'encapsulation

des tunnels n'utilise plus d'en-tête IP avec "loose source routing" mais le protocole "IP dans IP" (RFC1241). La plage d'adresses 224.0.0.0 à 224.0.0.255 a été réservée pour les protocoles de routage. Les tunnels établis entre démons mrouted constituent un réseau virtuel audessus de l'infrastructure et des domaines de routage monopoint qui ne voient passer que des paquets IP ordinaires...

Principales caractéristiques d'un tunnel mrouted : adresses d'extrémités, seuil de TTL (threshold) et métrique. Le seuil permet de délimiter des aires de propagation : pour qu'un datagramme multipoint puisse traverser un tunnel il faut que son champ TTL ait une valeur supérieure au seuil du tunnel. Par ailleurs son TTL est décrémenté de 1 à chaque traversée de mrouted. Le métrique représente le "poids" du tunnel et est utilisé comme critère de choix entre plusieurs tunnels (backup).

Pour fonctionner il est nécessaire que l'OS sur lequel mrouted s'exécute supporte le multipoint, ce qui n'est pas encore le cas de tous. Pour certains de ceux qui ne le supportent pas en standard (cas de SunOS par exemple) des extensions existent sous forme de patches et d'ajouts au noyau.

mrouted est une plate-forme d'expérimentation et d'évolution pour IGMP et DVMRP et évolue assez régulièrement en apportant des améliorations significatives. La version actuelle est 3.6. Attention aux vieilles versions (2.x) fournies avec certains OS : elles sont fortement déconseillées...

5 Le MBone

MBone, pour Multicast backBone, est le réseau virtuel de diffusion multipoint. Il est issu d'expérimentations de l'IETF, à San Diego, en 1992 visant à diffuser sur l'Internet le son puis la vidéo de ses réunions plénières. Ce réseau est constitué, pour l'instant, de machines exécutant mrouted qui tissent une toile d'araignée constituée de tunnels au-dessus de l'Internet (cf Figure 1). Cette situation est transitoire en attendant le support du multipoint dans les routeurs de l'Internet et le choix d'un protocole de routage approprié.

La mise en oeuvre du MBone est basée sur le volontariat, il faut le considérer comme un réseau encore expérimental ayant des "sautes d'humeur" : il n'est pas rare que de sévères anomalies se produisent écroulant, voire crashant, les machines mrouted et/ou saturant des liens. Ces anomalies ont diverses causes : vieilles versions de mrouted, interactions entre routeurs PIM et mrouted, utilisateurs d'applications audio/vidéo générant un trafic abusif,...

5.1 Le FMBone

C'est la section française du MBone. Démarré fin 1993, son déploiement est coordonné par Christian Donot pour le compte de l'association Aristote qui y diffuse ses séminaires. Le point d'entrée (routeur primaire) est une machine dédiée située sur un réseau d'EDF connecté au réseau régional RERIF à 34Mb/s. La topologie du FMBone (cf Figure 2) se calque en gros sur celle des plaques régionales avec un routeur secondaire par plaque. Ces routeurs secondaires sont chargés de redistribuer les flux du MBone sur leurs plaques respectives.

La liste de diffusion mbone-fr@inria.fr est utilisée pour les échanges d'information à propos du FMBone. Inscription : mbone-fr-request@inria.fr. Cette liste est (trop) active car elle reçoit les messages de la liste européenne elle-même recevant ceux de la liste mondiale...

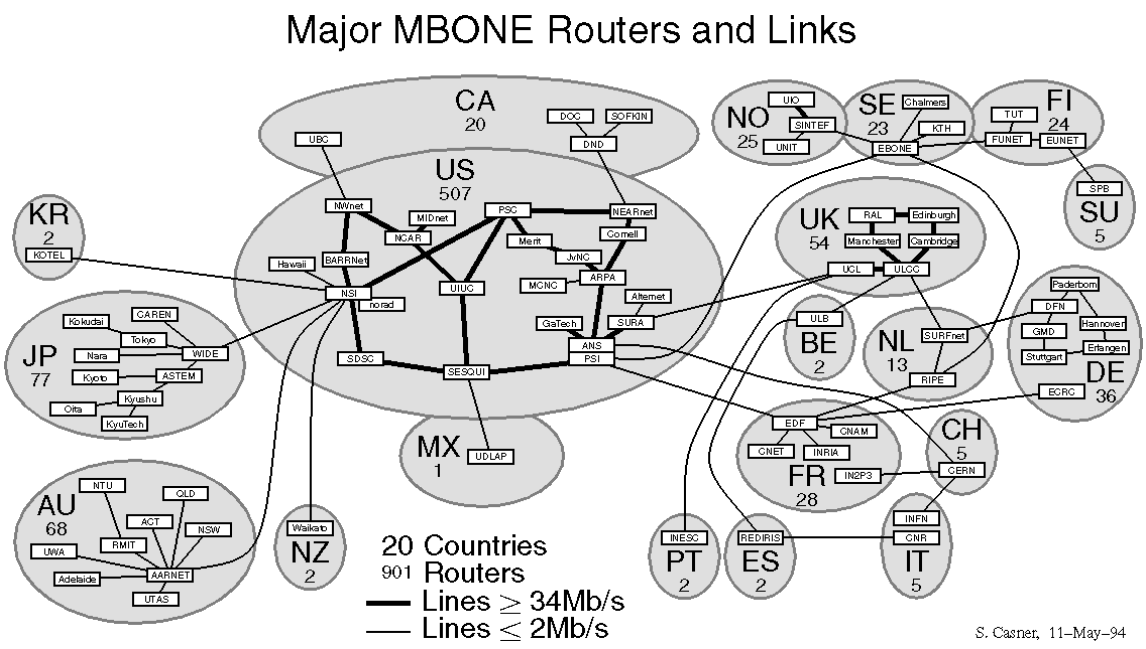

FIG. 1 – *Principaux liens du MBone*

5.2 Qu'est-ce qui circule sur le MBone ?

Outre les réunions annuelles de l'IETF ont y trouve de plus en plus de séminaires ou de conférences (par exemple, en France ceux d'Aristote), des images de satellites météorologiques en quasi temps réel, la retransmission des missions de la navette spatiale américaine (cf Figure 3), des "démonstrations" lors de salons,... jusqu'à des extraits de concert comme celui des Rolling Stones en novembre 1994 montrant l'intêret naissant d'entreprises commerciales pour l'utilisation du MBone.

Mais le MBone est également le support de diffusions à moins grande échelle: télé-réunions de travail, tableaux blancs partagés,... autrement dit les applications de travail collaboratif.

5.3 Comment se raccorder au MBone

L'utilisation d'applications de travail collaboratif à travers l'Internet ou le souhait d'assister aux séminaires diffusés sur le MBone suppose évidemment de s'y connecter. Pour un site "feuille" il faut :

- vérifier que l'on a de la bande passante en réserve : les applications de diffusion vidéo consomment facilement 100kb/s et plus par session. Un point d'accès à l'Internet à 512kb/s semble être le minimum requis...

- disposer d'un équipement implémentant la version de DVMRP de mrouted 3.6. S'il s'agit d'une station Unix vérifier que son système sait faire du multipoint (Solaris 2.3, Linux,

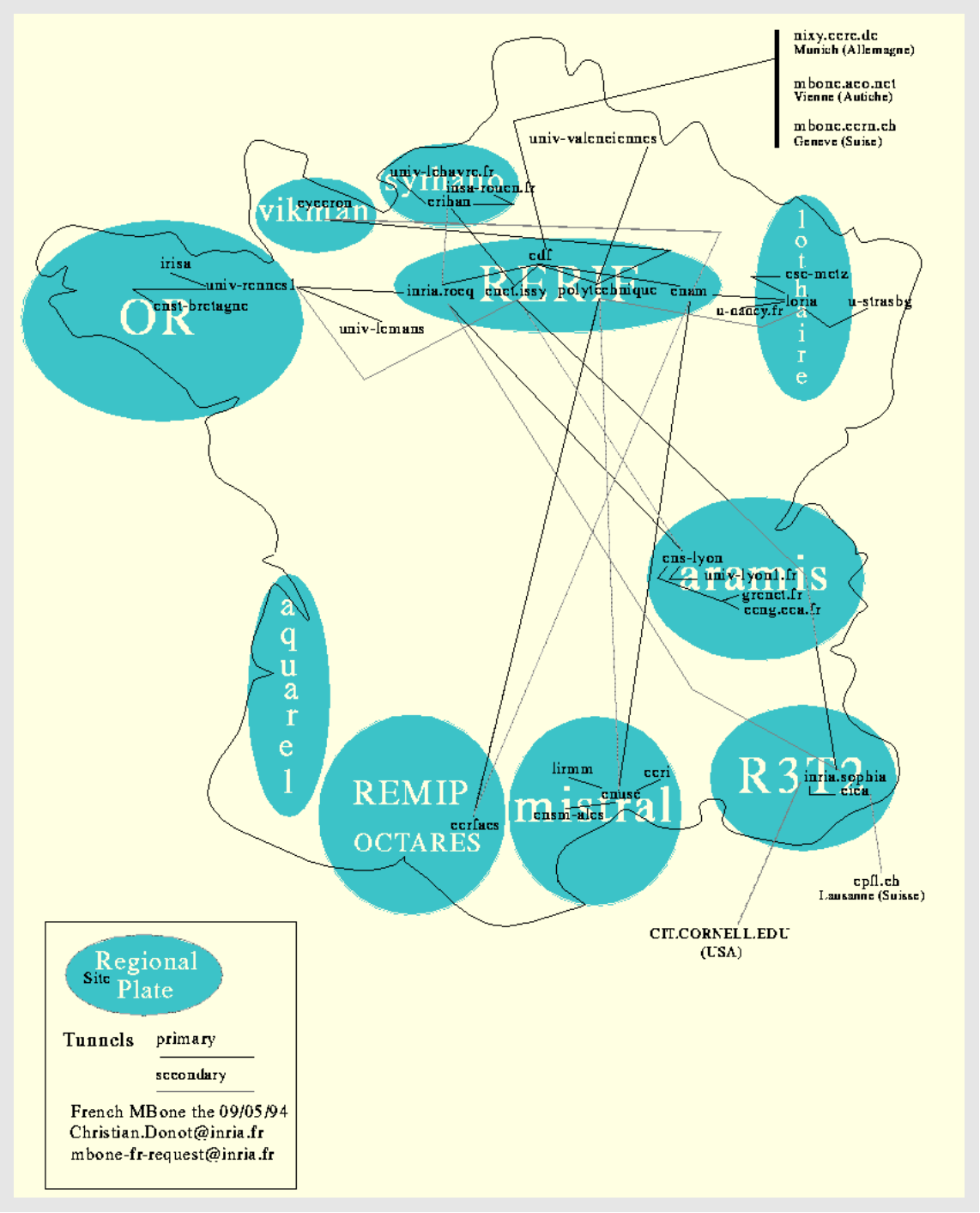

FIG. 2 – *Partie française du MBone*

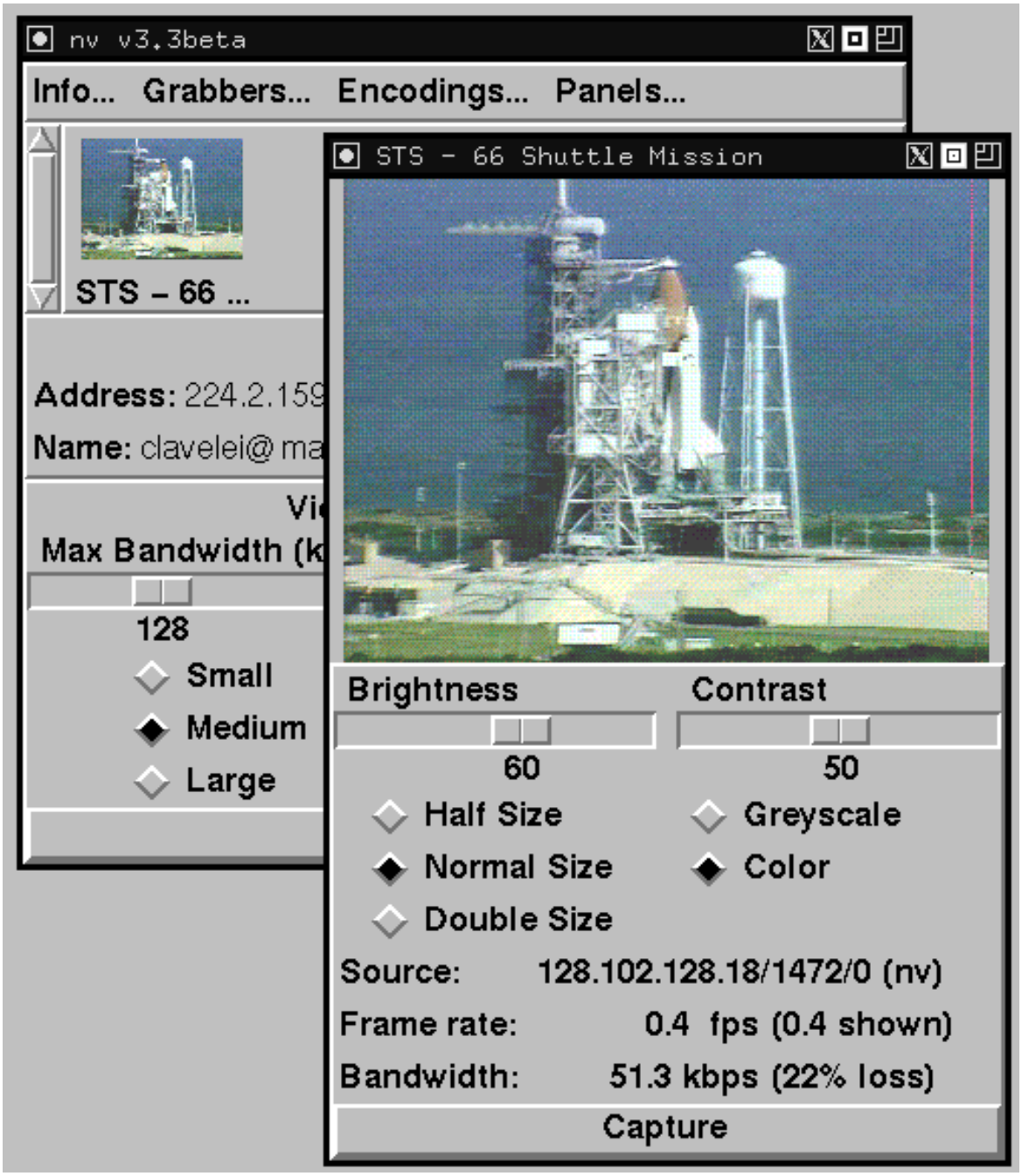

FIG. 3 – *Exemple de diffusion sur le MBone : la mission STS66 de la navette spatiale*

NetBSD, BSD4.4, IRIX 5.x,NextStep 3.3, OSF/1 2.0,...) ou peut le faire à l'aide d'extensions du domaine public (SunOS 4, AIX 3.2, IRIX 4.x, Ultrix) ou constructeur (HP-UX) et récupérer la dernière version de mrouted. Un routeur Cisco en version 10.x devrait être capable d'interopérer avec un routeur mrouted et constituer le point d'accès au MBone. Proteon dispose d'une (vieille) version de DVMRP. Il semble que 3Com ait annoncé la disponibilité de DVMRP 3.5 et MOSPF sur ses routeurs. Bay Networks supporte DVMRP 3.3.

– contacter le coordinateur du FMBone (Christian.Donot@inria.fr) et/ou utiliser la liste mbone-fr@inria.fr pour "se faire adopter" et déterminer quel est le point de "tunnelling" optimal. À cette occasion il peut être nécessaire de modifier la topologie existante en créant un nouveau noeud

– suivant le cas configurer le routeur ou le mrouted en coordination avec le gestionnaire de l'autre "bout du tunnel". Il est possible de limiter le débit sur un tunnel pour "limiter les dégâts" en cas de trafic élevé mais cela se fait évidemment au détriment du fonctionnement des applications concernées...

– activer autant de routeurs multipoint que de réseaux à alimenter sur le site

– surveiller la consommation de bande passante du trafic multipoint...

  Dans le cas d'un noeud de rediffusion il faut un accès à au moins 1Mb/s et une machine de routage suffisament dimensionnée pour supporter la charge...