



HAL
open science

**Rapport scientifique (d'étape) du projet deréseau
régional d'expérimentations avancées sur un réseau haut
débit ATM de Grenoble C3I2 (CEA, CNET, CNRS,
IMAG, INRIA)**

Jean-Luc Archimbaud

► **To cite this version:**

Jean-Luc Archimbaud. Rapport scientifique (d'étape) du projet deréseau régional d'expérimentations avancées sur un réseau haut débit ATM de Grenoble C3I2 (CEA, CNET, CNRS, IMAG, INRIA). 1998. hal-00561008

HAL Id: hal-00561008

<https://hal.science/hal-00561008>

Preprint submitted on 31 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rapport scientifique (d'étape) du projet de réseau régional d'expérimentations avancées sur un réseau haut débit ATM de Grenoble C3I2 (CEA, CNET, CNRS, IMAG, INRIA)

2 novembre 1998

Rédacteur principal : Jean-Luc Archimbaud CNRS/UREC coordonnateur du projet

Ce rapport scientifique est un **bilan d'étape des expérimentations à Grenoble**. Son but est d'établir un compte-rendu de la mise en place et un état des expérimentations mais aussi d'en tirer des leçons sur l'utilisation d'ATM, sur la gestion d'un tel projet et sur l'administration d'un réseau multi-partenaires.

Ce document rappelle les objectifs, l'historique et la coordination de ce projet (chapitre 1), décrit l'architecture physique et logique du réseau (chapitre 2), développe les problèmes techniques de mise en place (chapitre 3), décrit les expérimentations applicatives (chapitre 4), se pose la question sur le choix d'ATM en temps que solution technique pour un réseau métropolitain ou régional (chapitre 5) et enfin essaie de tirer des leçons sur la conduite d'un projet de réseau expérimental et l'administration d'un réseau de production multi-partenaires (chapitre 6).

Des informations sont aussi disponibles en ligne : <http://www.urec.fr/C3I2/>.

1. Objectifs, historique, coordination

Le projet, tel que décrit dans les dossiers initiaux présentés à la région Rhône-Alpes, visait à **installer un réseau à haut débit** entre des partenaires grenoblois de la recherche pour conduire des **expérimentations de protocoles** (ATM, IP, IPv6 ...), de **services** (vidéoconférences, ...) et des **applications avancées** issues des laboratoires de recherche de Grenoble (calcul distribué, espace de travail réparti, ...). Il devait aussi **préparer l'évolution des infrastructures de réseau de campus, métropolitaines et régionales (ARAMIS)** en permettant aux ingénieurs d'acquérir un savoir faire dans les nouvelles technologies et d'expérimenter certaines architectures et modèles d'administration des équipements. Ces objectifs n'ont pas été modifiés et on peut considérer qu'**ils ont été atteints**.

La volonté de mettre un place un réseau haut débit expérimental à Grenoble n'est pas récente. De nombreuses réflexions ont été menées depuis plusieurs années dans ce sens en particulier sous l'impulsion de l'IMAG. Ces idées ont été reprises dans le cadre de l'association GNI@, avant d'aboutir à C3I2. La genèse a été longue, certainement pour 2 raisons : la définition précise des partenaires intéressés et des expérimentations ainsi qu'une offre acceptable techniquement et financièrement du seul opérateur qui pouvait offrir un service réseau haut débit, France Télécom.

Initialement le projet s'est appelé **C2I2**, initiales des 4 partenaires de départ : **CEA** Grenoble (Commissariat à l'Energie Atomique), **CNRS** (Centre National de la Recherche Scientifique), **IMAG** (Institut d'Informatique et de Mathématiques Appliquées de Grenoble) avec ses 2 tutelles universitaires INPG et UJF, et l'**INRIA** Rhône-Alpes (Institut National de la Recherche en Informatique). En cours de définition, le **CNET** de Meylan (Centre National d'Etude des Télécommunications) a manifesté son intérêt pour ce réseau et s'est joint aux premiers organismes, donnant **C3I2**.

Voici chronologiquement quelques faits marquants :

- . Printemps 96 : définition du projet, préparation du dossier régional mi-lourd 96, premiers contacts avec l'opérateur France Télécom.
- . Automne 96 : concertation avec nos collègues de Lyon pour harmoniser nos 2 projets similaires présentés à la région. Accord de principe de France Télécom pour établir une proposition de service haut-débit entre 4 sites grenoblois.
- . Décembre 96 : accord pour une première tranche de financement régional (mi-lourd 96).
- . Hiver 96-97 : organisation des comités C2I2. Définition précise du service avec France Télécom. Arrivée du CNET dans le projet.
- . Printemps 97 : préparation d'un dossier régional complémentaire pour l'opération mi-lourd 97.
- . Eté 97 : consultation des différents fournisseurs d'équipements réseaux pour l'équipement des sites.
- . Automne 97 : choix des équipements de site. Signature des contrats avec France Télécom.

- . **Décembre 97 : ouverture du service France Télécom pour le réseau C3I2 pour une durée de 1 an.** Réception et installation des équipements de site. Accord pour la seconde tranche de financement régional (mi-lourd 97).
- . Mars 98 : connexion du réseau C3I2 à SAFIR (réseau expérimental ATM national, préfiguration de RENATER 2).
- . Avril 98 : ouverture du réseau aux expérimentations applicatives. Retard du à différentes causes décrites dans le chapitre 3
- . Juin 98 : bascule de certains trafics de production entre le domaine universitaire et l'INRIA Montbonnot sur C3I2, et entre Paris-Grenoble et Lyon-Grenoble sur SAFIR-C3I2.
- . Sept 98 : bascule du trafic de production entre le polygone CNRS et le domaine universitaire sur C3I2.
- . Oct 98 : bascule d'une partie du trafic de production entre le site Viallet (INPG, IMAG) et le domaine universitaire sur C3I2. Contacts avec France Télécom pour la poursuite éventuelle de C3I2 (le contrat se terminant le 30/11/98).

Cette liste montre **qu'un tel projet nécessite du temps, 2 an ½**, pour passer de la définition jusqu'au transport du trafic de production, preuve que le réseau offre un service stable; et que les étapes sont nombreuses mais incontournables.

Aucun moyen humain n'a été affecté à plein temps à ce projet. Chaque partenaire a désigné un ou plusieurs représentants qui ont travaillé à la mise en place du réseau et des expérimentations en plus de leurs tâches quotidiennes. Le projet a donc dès le départ nécessité une coordination, qui s'est voulu légère mais efficace, à la fois décisionnelle et technique. 3 comités ont été formés :

- . Un **comité de pilotage**, organe décisionnel, constitué de :
 - . Jean Luc Archimbaud Directeur Technique de l'UREC (Unité Réseaux du CNRS) coordonnateur du projet
 - . Daniel Bois Directeur du CNET Meylan
 - . Pierre Laforgue Directeur des moyens informatiques de l'IMAG, Directeur technique d'ARAMIS
 - . Jean Potier Responsable Informatique du CEA Grenoble
 - . Jean-Pierre Verjus Directeur de l'INRIA Rhône-Alpes
 - . Jacques Voiron Directeur de l'IMAG
- . Un **comité technique-partenaires** qui regroupe les chefs de projet de chaque partenaire, chargé de prendre les décisions techniques, composé de :
 - . Jean Luc Archimbaud (CNRS/UREC) coordinateur
 - . Raoul Dorge (CEA)
 - . Joseph Lecourt (CNET)
 - . Jean-Luc Parouty (IMAG)
 - . Luc Saccavini (INRIA)
- . Un **comité technique-points d'accès**, réunissant les ingénieurs chargés de l'administration des équipements télécom sur les différents sites :
 - . Site INRIA Montbonnot : Luc Saccavini (INRIA) et Jean-Pierre Auge (INRIA)
 - . Site Viallet : Claire Rubat du Merac (INPG) et Jean-Luc Parouty (IMAG)
 - . Site domaine universitaire : Christian Lenne (CICG), Jean-Luc Parouty (IMAG) et Jacques Eudes (UJF)
 - . Site polygone : Raoul Dorge (CEA) et Daniel Guéniche (CNRS)
 - . Site CNET Meylan : Joseph Lecourt (CNET) et Salvador Salas (CNET)

Les 2 derniers comités sont très actifs et se réunissent ensemble, en moyenne toutes les trois semaines. Il faut remercier l'ensemble de ces 2 comités pour avoir « jouer le jeu » des expérimentations, leurs consacrant beaucoup de temps en plus de leur travail quotidien d'exploitation.

2. Architecture du réseau

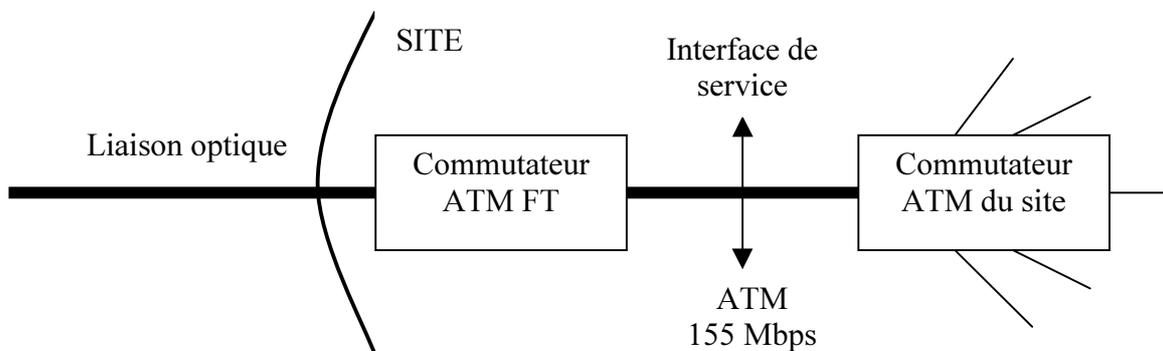
Les 5 partenaires du projet sont situés sur 5 pôles différents de Grenoble : le domaine universitaire (avec des laboratoires de l'IMAG, du CNRS, des universités, ...), les bâtiments de l'INPG avenue Félix Viallet (IMAG, CNRS, ...), le polygone scientifique (CEA, CNRS, ...), Montbonnot (INRIA) et Meylan (CNET).

Il faut d'abord souligner que 4 des 5 partenaires sont interconnectés, pour leur trafic de production, sur le réseau régional ARAMIS, lui même connecté à RENATER (donc à l'Internet) avec des prises en moyenne à 3 Mbps. Durant les expérimentations C3I2, ARAMIS a permis d'accéder à distance aux équipements C3I2 même lorsque l'accessibilité était impossible par le réseau C3I2 et a permis aussi de tester les protocoles de routages dynamiques de IP (cf expérimentations).

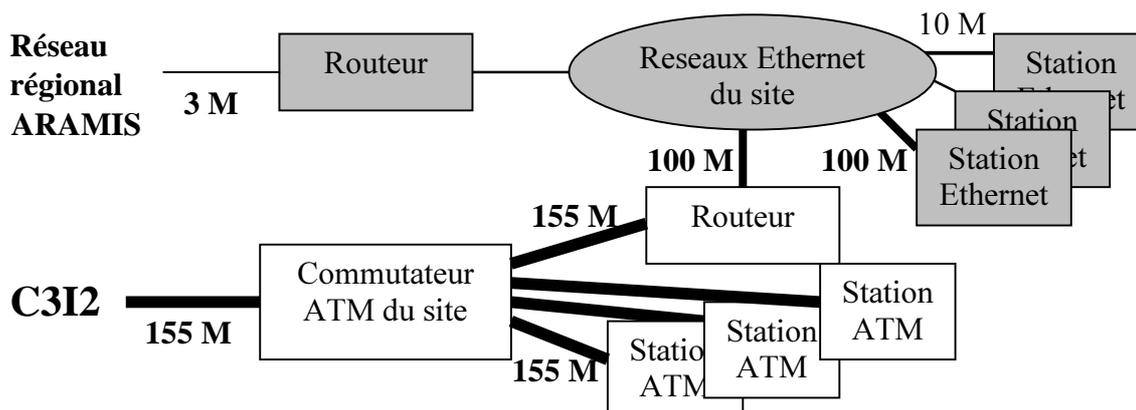
Les architectures décrites ci-dessous nécessitent, pour leur compréhension, une petite connaissance d'ATM. Elles permettent de montrer les nombreuses possibilités d'architectures que l'on peut construire sur une infrastructure ATM.

Architecture physique

Le réseau C3I2 est bâti sur l'offre de **VP (Virtual Path) ATM en mode CBR (débit constant) de l'opérateur France Télécom pour interconnecter les 5 pôles**. Actuellement ce service porte le nom d'**OMA (Offre Multiservices sur ATM)**. Chaque site est desservi par une liaison optique avec un commutateur ATM, équipement d'extrémité fourni et administré par l'opérateur, auquel le site n'a pas accès. De son commutateur l'opérateur offre au client une interface de service, ATM à 155 Mbps avec un media optique.



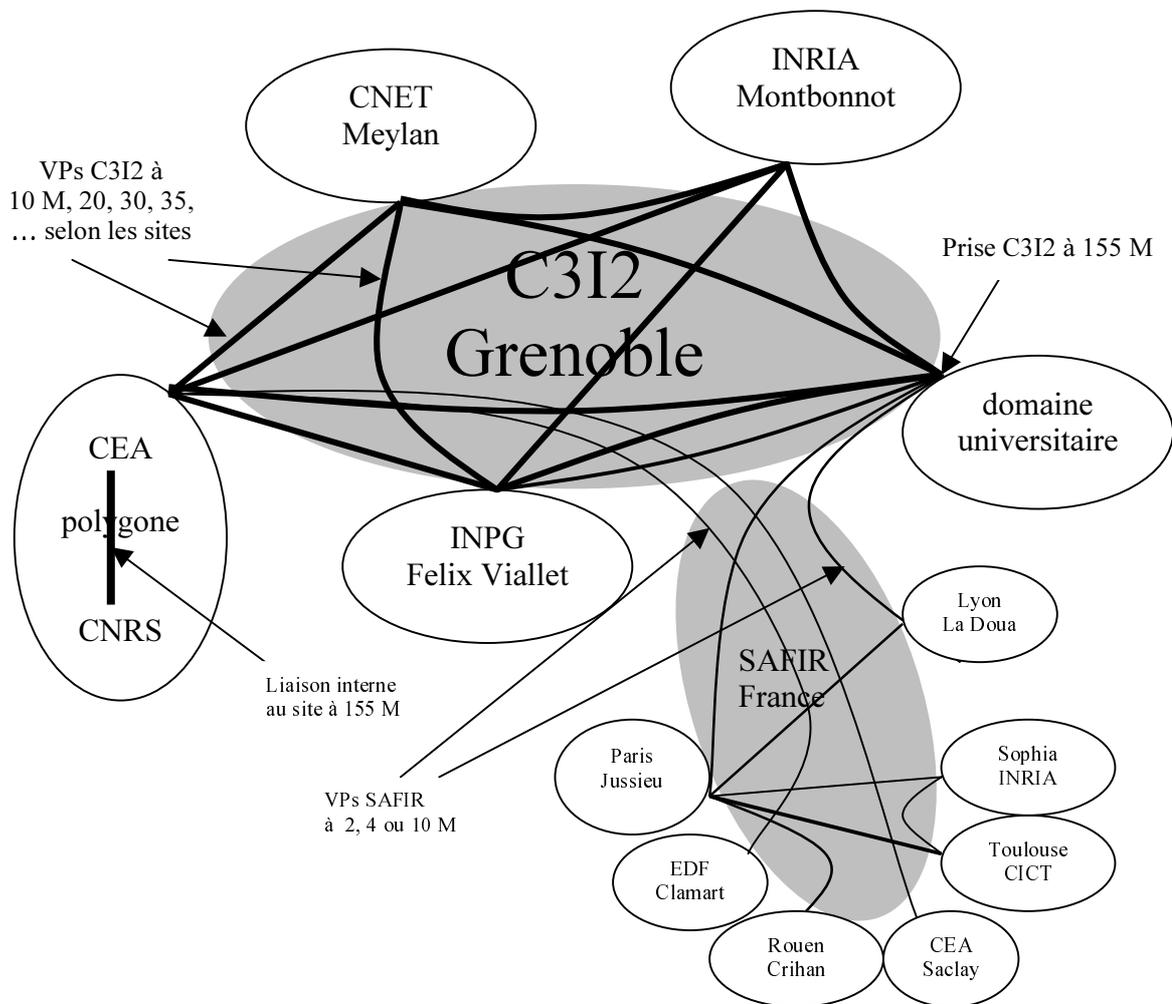
Sur cette interface de service le site connecte son propre commutateur ATM avec des stations ATM et un **routeur ATM-Ethernet permettant d'accéder au réseau de production du site** (en grisé sur le schéma suivant). Tous les sites ont en interne un réseau de production de type campus, Ethernet commuté à 100 M et parfois FDDI.



Chaque site dispose ainsi d'une bande passante de 155 Mbps vers le réseau C3I2. Celle-ci est découpée en VPs, l'équivalent de liaisons spécialisées à débit fixe et garanti, vers les autres sites.

L'architecture de réseau de VPs que nous avons construite est maillée. Chaque site est connecté aux quatre autres par au moins un VP. Les changements de configuration des VPs (nouveau VP, modifications de débit des VPs existants, ...) peuvent être demandés par télécopie à l'opérateur et sont pris en compte dans les 24 heures, ce temps s'étant avéré largement suffisant comme délai de réaction.

Cette infrastructure est connectée au réseau national expérimental ATM SAFIR (<http://www.renater.fr/Safir/safir.html>), préfiguration de RENATER 2 qui raccorde actuellement Grenoble, Lyon, Paris, Toulouse, Rouen et Sophia avec un système de VPs similaire à C3I2. Concrètement, la liaison optique C3I2 qui dessert le domaine universitaire connecte aussi celui-ci à l'infrastructure SAFIR, et à travers lui tous les sites de C3I2. Du domaine universitaire partent 2 VPs à 10 M, l'un vers Paris, l'autre vers Lyon. Ces 2x10 M sont déduits des 155 de la prise C3I2. Toujours dans le cadre de SAFIR, le CEA Polygone a 2 VPs à 2 et 4 M vers l'EDF Clamart et le CEA Saclay. Sur les sites ont été installés, entre autres, 4 commutateurs ATM LS1010 de CISCO, 2 commutateurs ASX de FORE, 4 routeurs CISCO avec une carte ATM, des commutateurs ATM-Ethernet de MADGE, ainsi que différentes stations souvent multimédia ATM ou Ethernet : PC, SUN, DEC, SGI, ...



L'architecture logique ATM

La majorité des applications qui utilisent C3I2 sont écrites sur une pile logicielle TCP/IP. Beaucoup, qui initialement désiraient de l'ATM natif ont, pendant la mise en place du réseau, été portées sur TCP/IP. Néanmoins une connectivité uniquement IP n'était pas envisageable pour supporter certaines applications qui étaient prévues (routage distribué, interconnexion de PABX, établissement de chemins permanents pour IPv6, multicast national) mais aussi pour pouvoir facilement isoler différents flux.

Ainsi le but a été de **pouvoir établir une connexion ATM directe entre toutes les stations (ou routeurs ATM) du réseau C3I2, ainsi que du réseau SAFIR**, sans passer par un routeur IP, élément qui est un goulot d'étranglement pour les performances.

Sur l'infrastructure de VPs décrite précédemment et fournie par l'opérateur, nous avons construit notre propre réseau privé ATM par transport de notre signalisation à l'intérieur des VPs de l'opérateur ("**tunneling**"), seule méthode pour disposer d'un réseau ATM « dynamique ». Ainsi, lorsque l'on établit une connexion ATM (concrètement un VC, Virtual Channel) entre 2 stations C3I2, les équipements de France Télécom sont totalement transparents et ce sont les commutateurs ATM des sites qui gèrent les différents protocoles de signalisation ATM.

Nous avons suivi le **plan d'adressage ATM national de SAFIR** (aussi appelé plan d'adressage RENATER 2) où un intervalle d'adresses est réservé à Grenoble. Les 10 octets 39.25.0f.00.00.00.2d.00.79.01 constituent le préfixe attribué à C3I2 par SAFIR. Nous avons ensuite attribué à chaque site de C3I2 une plage d'adresses dans cet intervalle. Ainsi le domaine universitaire a eu le préfixe sur 11 octets 39.25.0f.00.00.00.2d.00.79.01.01, l'INRIA 39.25.0f.00.00.00.2d.00.79.01.02, ...

Chaque site a pu ensuite utiliser les octets restants pour numéroter ses stations (une adresse ATM fait 20 octets où uniquement les 13 premiers sont configurables). Ainsi l'adressage est unique sur C3I2 mais aussi sur SAFIR et toute station ATM de C3I2 peut atteindre une autre station de C3I2 mais aussi de SAFIR directement en ATM (sans passer par un routeur).

Même si on considère qu'il est obligatoire de segment un réseau national de production par des routeurs IP entre les régions, pour avoir une architecture avec plusieurs réseaux IP, la possibilité d'avoir un chemin direct ATM entre toutes les stations en France sur ce réseau expérimental laisse une très grande liberté d'architecture. La connexion directe ATM est toujours plus efficace, en performance que le passage par un routeur. Suivant les applications nous utilisons soit le chemin direct ATM, soit le chemin avec passage par un routeur IP.

Routage ATM

Sur C3I2 nous utilisons le protocole de routage dynamique **PNNI** (Private Network to Network Interface) entre les 4 commutateurs CISCO et **IISP** (Interim Inter Switch Protocol) avec un routage statique donc avec les commutateurs FORE (cf le chapitre problèmes de mise en place) et avec les équipements hors Grenoble de SAFIR. PNNI, qui évite une configuration manuelle et fastidieuse des « routes » dans chaque équipement semble très bien fonctionner avec des équipements homogènes (CISCO chez nous) avec un nombre restreint de nœuds, sans hiérarchie (nous n'avons pas testé le multi-niveaux inutile dans notre configuration).

PNNI permet de tirer tous les bénéfices de l'architecture maillée. Si le VP direct entre 2 sites est « coupé » ou saturé, le trafic peut faire un crochet par un troisième site intermédiaire.

Architecture LANE (LAN Emulation)

Sur l'infrastructure C3I2 et SAFIR, nous avons construit l'équivalent de **2 réseaux Ethernet (LANE)**, l'un métropolitain C3I2 et l'autre national. Cette architecture permet de raccrocher des stations derrière des commutateurs ATM-Ethernet (edge device) et aussi de disposer de la fonction basique de diffusion d'Ethernet qui peut être utile pour les applications multicast.

Le commutateur CISCO sur le domaine universitaire est serveur LECS et LES pour le LANE C3I2. Les serveurs pour le LANE SAFIR sont à Jussieu.

Architecture CLIP (Classical IP)

De la même manière que LANE nous avons bâti sur C3I2 une architecture logique CLIP, avec un sous-réseau IP pour les équipements de communications (routeurs, ...) et un sous-réseau IP pour les stations. L'idée est de séparer logiquement les équipements d'interconnexion des stations utilisateurs. Le routeur CISCO du domaine universitaire est serveur ARP pour le CLIP C3I2. Il existe aussi un réseau CLIP SAFIR ou le serveur ARM est installé à Jussieu. Nous pouvons aussi enregistrer une station C3I2 dans le réseau CLIP de SAFIR.

Adressage IP

Sur C3I2, nous utilisons un réseau de classe C officiel, segmenté en 4 sous-réseaux : CLIP et LANE pour les équipements réseaux (commutateurs, routeurs) et séparément pour les stations. Nous avons aussi une plage d'adresse réservée dans les 2 réseaux de classe C du CLIP et du LANE nationaux de SAFIR.

Routage IP

Le routeur ATM du domaine universitaire assure le routage entre les sous-réseaux IP de C3I2 (LANE et CLIP) et avec les réseaux SAFIR. Le réseau IP de C3I2 est annoncé sur l'Internet par le routeur du domaine universitaire connecté sur ARAMIS. Au cœur de C3I2 nous utilisons le routage IP statique avec une route par défaut, mais pour transporter du trafic de production nous utilisons BGP (cf chapitre expérimentations).

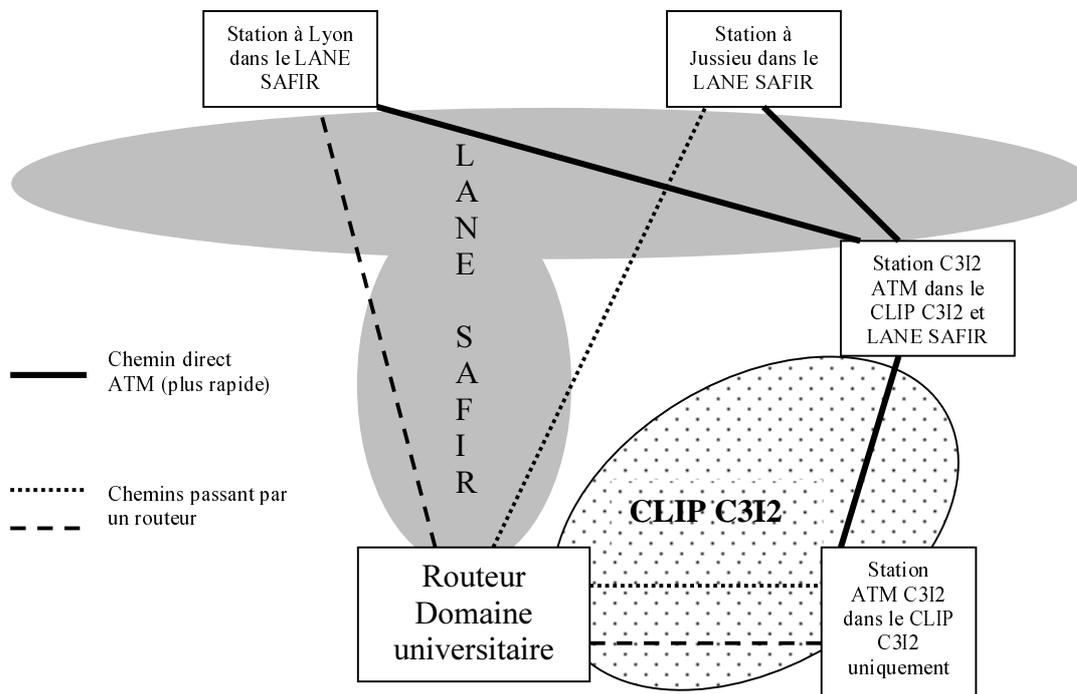
Connexion des stations

Initialement, comme le voudrait une bonne méthode expérimentale, nous désirions séparer physiquement les réseaux de production et ce réseau expérimental. Mais les sites n'ayant pas en interne un réseau ATM, pour connecter des stations d'expérimentation en ATM il fallait une fibre optique entre la station et le commutateur d'entrée, denrée rare sur les sites. Donc peu de stations sont directement connectées sur ce réseau expérimental. Beaucoup sont sur le réseau Ethernet de production et nous utilisons le routage statique IP par station dans les routeurs sur le chemin pour orienter le trafic vers C3I2 ou vers ARAMIS. Avec ce type de raccordement, nous vérifions le

débit disponible sur l'ensemble du chemin en particulier sur les réseaux de production, pour garantir un minimum de bande passante.

Avec ces différentes architectures, nous avons toutes les variantes possibles de connectivité avec ATM, en ajustant simplement la configuration des stations. Ainsi une station ATM (ou un routeur C3I2) peut appartenir au réseau CLIP de C3I2, au réseau CLIP de SAFIR, au LANE de C3I2 ou au LANE de SAFIR, à plusieurs de ces réseaux, voire à tous. Le fait d'appartenir par exemple au LANE SAFIR permet de ne pas avoir à traverser un routeur pour accéder aux stations de SAFIR.

Exemple de connectivité :



Administration

Chaque ingénieur de site administre son routeur, son commutateur et ses stations. Les réunions régulières des comités techniques permettent de se coordonner et plusieurs listes de diffusion électroniques sont très actives. Un outil très simple (à base de ping avec petits et gros paquets et traceroute) vérifie régulièrement, toutes les 5 minutes, l'accessibilité IP des équipements télécom et de certaines stations. En cas d'anomalie (accessibilité ou routage), un message électronique est envoyé (à travers ARAMIS évidemment) aux administrateurs de sites concernés. L'outil MRTG avec un accès Web installé sur le site INRIA permet d'avoir des relevés de trafic et de charge de tous les commutateurs et routeurs de C3I2.

3. Les problèmes de mise en place

Il y a eu de très nombreux problèmes de mise en place, phénomène normal et prévisible vues les nouvelles technologies utilisées et les versions expérimentales des logiciels fournis par les constructeurs sur les équipements de communication. Ces erreurs ont malheureusement beaucoup retardées l'ouverture du réseau aux expérimentations applicatives,

Ce chapitre fait un court résumé de chaque problème et essaie d'en tirer une leçon pour la mise en place d'un réseau, au risque de formuler des lapalissades.

Où 155 Mbps théoriques deviennent 134 Mbps utiles

Notre contrat auprès de France Télécom stipule un accès à 155 Mbps. Dans un premier temps, nous avons réfléchi à une architecture avec des VPs à hauteur de 155 Mbps par site. Mais lorsque nous avons demandé à l'opérateur de la mettre en place, nous avons eu les précisions sur ce qu'indique 155 M., C'est le débit physique de l'accès. ATM n'est pas vraiment le protocole utilisé directement sur la fibre optique, c'est le protocole SDH (Synchronous Data Hierarchy) qui transporte les cellules ATM. Donc aux 155 Mbps, il faut ôter la bande passante utilisée par les entêtes SDH, ce qui ramène la bande passante à 149 Mbps. D'autre part, une règle d'ingénierie retenue dans le réseau ATM de France Télécom, impose de réserver une marge de manœuvre de 10 % de cette bande passante pour l'exploitation et la bonne marche des commutateurs de l'opérateur. Il a donc resté 134 M bps par site pour définir les différents VPs.

Sur tous les types de réseau, il est obligatoire de connaître le débit utile offert par le cœur du réseau et aux points d'accès, et pas uniquement le débit théorique qui est souvent physique.

Shaping (lissage)

Lorsque l'on crée plusieurs VPs (entités logiques) avec un débit fixé, maximal, sur une interface physique à 155 Mbps, comme c'est le cas des accès C3I2, il est nécessaire que ces maxima de débit soient respectés pour chaque VP individuellement. Il ne faut pas émettre plus de 30 Mbps dans un VP à 30 Mbps, même si le coupleur permet d'écouler 155 Mbps. Pour faire respecter ce contrat, comme tout opérateur, France Télécom installe une fonction de contrôle (policing) sur ses équipements qui élimine les cellules ATM au dessus de la bande passante de chaque VP. Sur le site, il faut que le commutateur implémente une fonction de lissage (shaping) du trafic qui permette d'absorber les rafales venant des autres interfaces à 155 Mbps dans des buffers pour les émettre avec un certain espacement de manière à arriver à un débit inférieur au débit du VP. Dans notre cas, cette fonction est appelée shaping multi-VPs sur un même port physique. Cette fonctionnalité a été un critère de choix pour nos commutateurs.

Malheureusement, malgré les promesses du constructeur et la documentation, les premières versions de logiciel livrées assuraient incorrectement cette fonction de shaping. Nous avons mis du temps à nous en rendre compte et à attribuer des mauvais fonctionnements à cette déficience. Par exemple, les datagrammes pings standards (de 64 octets) étaient transportés correctement sur le réseau mais pas les datagrammes de 1000 octets. La raison était que la fonction de contrôle de notre opérateur accepte des rafales de trois cellules (une cellule contient 48 octets de données utiles) au dessus du

débit du VP, mais pas plus. Les « pings standards » étaient donc acceptés par le réseau mais pas les pings de plus de 48 x 3 octets. Un mauvais shaping se manifeste principalement par le fait que des petits paquets IP sont transférés sans problème alors que les gros ne passent pas.

Avec l'aide de France Télécom qui, en temps réel, peut nous indiquer si des cellules ATM sont rejetées par ses équipements à cause d'un mauvais shaping de notre part, au bout d'un certain temps, nous avons pu mettre au point une méthode pour vérifier que cette fonctionnalité de lissage était correctement assurée par nos équipements.

Il est très difficile de diagnostiquer un mauvais shaping dans un équipement ATM et plus généralement de reconnaître une mauvaise fonction de base (physique) dans un équipement réseau.

Utilisation du VP 0 et standardisation

Pour les expérimentations, nous avons installé des équipements de routage distribué Ipsilon sur 2 sites distants avec un commutateur ATM Ipsilon réservé à ces équipements, sur un des sites. Ces routeurs Ipsilon nécessitent un « tunneling » de VP entre les commutateurs CISCO d'entrée de site. Ils utilisent le VP 0 de manière tout à fait propriétaire, sans rapport avec les standards habituels de l'ATM Forum qui réservent ce VP pour la signalisation UNI, Il a fallu déclarer sur les commutateurs CISCO un paramétrage non documenté, que uniquement les développeurs de CISCO ont été capables de nous fournir. Lorsque l'on a essayé de savoir qui de CISCO ou de Ipsilon ne respectait pas les standards, on a eu un débat de spécialistes sans arriver à une conclusion claire pour un utilisateur. En simplifiant Ipsilon indiquait que rien dans les standards ATM Forum n'obligeait à réserver le VP 0 pour la signalisation, ce qui semble exact, et CISCO que tout le monde utilisait le VP 0 pour cette fonction, ce qui est aussi exact (sauf le constructeur Ipsilon). CISCO a montré de la bonne volonté pour résoudre ce problème car dans notre cahier des charges demandait l'interconnexion de routeurs Ipsilon.

La leçon semble double : **il est préférable d'implémenter les standards comme « le plus grand nombre » et il est très difficile de prouver qu'un constructeur ne respecte pas les standards.**

PNNI

PNNI (Private Network to Network Interface) est un protocole de routage dynamique sur ATM. Il fonctionne parfaitement entre les commutateurs CISCO, par contre entre CISCO et FORE il n'est actuellement pas interopérable. Il a été très difficile de diagnostiquer ce problème car les tables de routage ATM sur les commutateurs étaient correctement mis à jour, mais certaines applications entre les « sites équipés de commutateurs FORE et les autres de CISCO fonctionnaient à certains moments et pas dans d'autres, de « manière un peu aléatoire », apparemment sans logique. Le problème vient d'un refus d'ouverture de circuit virtuel, qui demande une signalisation PNNI, dans le sens CISCO vers FORE, alors que dans l'autre sens elle est acceptée. Ainsi lorsqu'un CV bidirectionnel était déjà ouvert dans le sens FORE vers CISCO, le trafic qui pouvait utiliser ce CV était écoulé sans problème. S'il y avait besoin d'une ouverture de CV dans le sens CISCO vers FORE (les CV se ferment au bout d'un certain temps d'inactivité) alors la connexion ne s'établissait pas et les données n'étaient pas transmises.

Le problème est dû à un bug CISCO qui positionne mal un « information element » dans le CALL SETUP d'ouverture de CV lors de l'utilisation de VP différent de zéro pour la signalisation PNNI, ce qui est notre cas (avec le tunneling de VP), cas rare dans le monde. Actuellement nous attendons le code correcteur de CISCO et avons contourné le problème par du routage statique (IISP) entre CISCO et FORE.

La leçon est qu'une très bonne connaissance des standards et un analyseur ATM sont obligatoires pour résoudre ce type de problèmes.

Performances décevantes avec le super ordinateur Cray T3E

Plusieurs expérimentations nécessitent l'accès à un super ordinateur Cray T3E équipé d'une carte ATM. Dans notre cas, une station à distance d'abord connectée sur un réseau Ethernet 10 M avait de très bons résultats lors des transferts avec le T3E via C3I2. Quand nous l'avons connectée sur un Ethernet à 100 M, pensant augmenter les débits de transfert obtenus, les résultats sont devenus catastrophiques, bien inférieurs aux précédents. D'autre part, les tests de performance effectués localement, sans transiter par le réseau C3I2, montraient des résultats convenables. Longtemps, nous avons pensé à un problème de shaping dans le réseau, ce qui n'était pas le cas. Une fois encore, il a fallu avoir recours à un outil d'analyse ATM pour diagnostiquer ce dysfonctionnement. Les communications expérimentales sur le Cray T3E traversant plusieurs routeurs, la taille des paquets de données IP est beaucoup plus petite que lors d'une connexion directe, pour se prémunir de toute fragmentation intermédiaire (MTU de 512 octets contre 9180). La solution pour contourner ce problème aurait été d'établir une connexion directe avec le superordinateur Cray T3E, mais cette configuration allait à l'encontre de la politique de sécurité du CEA en matière de communications informatiques.

Une démarche logique pour éliminer un élément défaillant ne peut se substituer à des outils d'analyse.

Mauvais fonctionnement des couches TCP/IP sur les OS

Nous avons aussi mis en cause le réseau ATM de manière injustifiée dans un autre cas. Des tests de ttcp (outil de mesure de performance TCP) entre 2 stations de 2 sites distants, donnaient des résultats corrects lorsqu'on effectuait un transfert. Mais lorsqu'on lançait deux transferts en parallèle la somme des débits obtenus était loin des performances précédentes dans un sens, alors que dans l'autre les débits obtenus étaient normaux. En fait, des deux côtés, les stations n'étaient pas identiques en terme d'Operating System. Dans un des OS les couches TCP/IP savaient mal gérer 2 sessions TCP en parallèles. Moralité : **il faut avoir des stations identiques en terme de matériel et de logiciel pour faire des tests de performance et valider une infrastructure de réseau.**

Interprétation de mauvaises performances TCP

Au fil des tests de performances de TCP sur C3I2, il est apparu clairement qu'il faut être très prudent dans l'interprétation des résultats et hormis la qualité des logiciels sur les stations et les performances des stations il faut aussi tenir compte des équipements traversés qui peuvent imposer des tailles de segments plus ou moins grands ou fragmenter les segments, et du paramétrage de TCP sur les stations comme la taille des fenêtres d'émission et de réception (qui sur certains OS demande de refaire un noyau).

Moralité : **il faut tenir compte de l'ensemble de la chaîne dans des mesures de performance.**

4. Les expérimentations

Ce chapitre décrit l'état des expérimentations à la date d'aujourd'hui.

Travail collaboratif : mediaspace

Mediaspace est un outil de travail collaboratif en environnement géographiquement réparti qui vise à créer un espace de travail virtuel, avec un rapprochement visuel au moyen de caméras "de bout de couloir" et de bureaux, qui donnent autant de fenêtres sur l'écran des stations des membres du groupe. Cet outil utilise la fonction multicast IP. Sur C3I2 il fonctionne entre 2 sites, connectés par un tunnel IP au dessus d'ATM. Le débit disponible permet d'avoir une très bonne définition ainsi que la possibilité d'utiliser l'audio dans les communications entre membres du groupe.

Mediaspace est développé par l'équipe IIBM du laboratoire CLIPS de la fédération IMAG.

Réalité virtuelle

Cette plate-forme formée de plusieurs stations et caméras connectées par un commutateur ATM crée un espace de réalité virtuelle avec un système de son spécialisé. Tournant déjà en local, l'expérimentation vise à éclater cet espace sur 2 sites avec une connexion directe ATM entre les 2 commutateurs ATM de 2 plates-formes.

Cette plate-forme est développée par l'équipe VIS du laboratoire GRAVIR (IMAG)..

CAONET

CAONET est une plate-forme de CAO partagée en réseau et destinée à accueillir des projets de conception de systèmes et de circuits, avec une mise en commun des logiciels, ainsi qu'un partage de bases de données, de bibliothèques de cellules et de macrofonctions (IPs). La conception de circuits nécessitant un affichage de très bonne définition, le réseau doit transporter un gros volume de données en un temps très court, pour un travail interactif. Cette plate-forme fonctionne entre 3 sites de C3I2. Les équipes impliquées sont au CNET et dans laboratoire LEMO (CNRS et INPG).

Post-traitement de résultats de calcul à la volée

Il est nécessaire de faire du post-traitement à la volée de résultats de simulation d'écoulements entre 3 stations sur un site et un T3E sur un autre site avec utilisation de PVM et des volumes de données de l'ordre de 50 Moctets à transférer en moins de 4 minutes à chaque pas. C3I2 permet de mettre en œuvre cette utilisation, impossible sur le réseau de production. Les calculs sont effectués par l'équipe MOST du LEGI (CNRS, INPG, UJF) sur le calculateur du CEA..

Calcul distribué

Cette expérimentation vise à effectuer entre plusieurs calculateurs répartis sur 2 sites, de la programmation de machines parallèles pour le calcul haute performance, de la modélisation sur plates-formes de calculs distribués, de la visualisation et ajustement en temps réel de modélisations 3-D et des échanges importants de fichiers de résultats (128 M à 1 Goctet par pas de temps). Les partenaires impliqués sont le CEA, l'Observatoire de Grenoble, l'IMAG et l'INRIA.

Expériences de télé-enseignement

Il a été réalisé plusieurs séances de télé-enseignement. Un exemple a été un cours d'apprentissage d'un logiciel flux-expert, logiciel de simulation de phénomènes physiques, entre un professeur dans le laboratoire Génie Atomique sur un site et plusieurs élèves de l'INPG sur un autre site avec l'utilisation de PC et de stations de travail HP et SUN multimedia ainsi que les logiciels du MBONE.

Séances de travaux pratiques répartis

Une école d'ingénieurs en électronique et en télécommunication (ENSERG) est répartie sur 2 sites. Elle désire utiliser l'infrastructure C3I2 pour monter des séances de travaux pratiques sur ATM et l'organisation d'autres séances de TPs (sur divers sujets) répartis sur 2 sites.

Téléconférences

De nombreuses télé-réunions et téléconférences ont été réalisées sur C3I2 avec différents outils tels que IP/TV, Netmeeting, les outils du MBONE ... C3I2 permet de tester ces outils sur un réseau avec une bonne qualité de service (bande passante, délai, gigue) non disponible sur le réseau de production. Ainsi de nombreux laboratoires se sont équipés de stations multimédia avec ces outils pour une utilisation « en vraie grandeur », professionnelle, à travers C3I2. Généralement ces outils utilisent le protocole IP multicast.

IPv6

Un réseau natif IPv6 a été créé, métropolitain et national (via SAFIR) avec des routeurs et des stations IPv6 en faisant tourner des codes spécifiques pour, en particulier, tester le routage sur IPv6. Techniquement ceci est réalisé à base de PVC ATM, démontrant que l'on peut construire un réseau « physique » indépendant sur un réseau général ATM.

Interconnexion de PABX

Des tests d'interconnexion d'autocommutateurs téléphoniques (PABX) sont prévus sur ATM entre 3 sites, soit directement avec une carte ATM dans le PABX ou avec une carte G703 dans le commutateur ATM.

Trafic d'exploitation – routage dynamique BGP4

Une partie du trafic de production entre les sites de C3I2 (et aussi de SAFIR) passe sur le réseau expérimental avec un back up sur le réseau standard de production ARAMIS-RENATER en cas de problème. Hormis le fait que cette bascule démontre que le réseau

ATM est stable et fiable, cela a permis de tester le protocole de routage dynamique BGP et le routage IP sur des réseaux maillés. Chaque site a ainsi mis en place un routage dynamique BGP pour utiliser les 2 voies de sortie possibles. Certains problèmes ont été soulevés avec BGP : le choix d'une route par rapport à l'agrégation des chemins, et le temps de convergence : pour l'instant nous n'avons pas pu faire descendre le temps de convergence en dessous de 2 mn (intervalle des hello BGP) car l'information de rupture de CV ATM ne remonte pas à BGP.

5. ATM était il un bon choix ?

Le but initial de C3I2 était d'avoir une infrastructure métropolitaine haut-débit sur lequel nous aurions pu tester différentes technologies de réseau. Pour ce faire nous aurions désiré disposer de fibres optiques réservées, entre chaque site de Grenoble. Le seul opérateur, France Télécom n'a pas voulu nous fournir ces liens et nous a proposé son offre ATM. Nous avons négocié pour avoir le minimum de service assuré par l'opérateur (pas de routeur opérateur par exemple) pour pouvoir avoir le maximum de liberté de tests dans les protocoles réseau. C'est ainsi que nous avons abouti sur **l'offre VP ATM de France Télécom qui était et est encore la seule possibilité pour avoir de tels débits à Grenoble, donc ATM était incontournable.** C'est un service ATM de base, un VP étant simplement l'équivalent d'une liaison spécialisée point à point, cela nous laissant la possibilité de construire autant de réseaux logiques que nécessaires. Ce service minimal s'est avéré un bon choix pour valider des solutions ATM et mener des expérimentations.

Sur C3I2, **le service ATM est très peu utilisé de bout en bout**, c'est à dire de station à station. En effet, les cartes ATM sont très chères par rapport à l'équivalent en Ethernet 100 M. De plus, les sites ne sont pas équipés de réseau ATM interne donc la connexion d'une station nécessite une fibre optique dédiée entre le commutateur ATM d'entrée et la station souvent dans un autre bâtiment. Beaucoup d'expérimentations qui au départ désiraient de l'ATM natif (car ATM était le synonyme de haut débit) se sont révélées avoir uniquement besoin de bande passante et fonctionner parfaitement sur IP, via Ethernet puis ATM, avec la bande passante désirée. Ainsi, nous n'avons pas pu utiliser les fonctionnalités d'ATM qui ont trait à la qualité de service (hormis la bande passante garantie). Cela nécessite des applications en ATM natif, mais aussi des équipements ATM qui peuvent implémenter les mécanismes nécessaires pour supporter ces fonctionnalités sur des VC, ce qui ne semble pas être disponible.

Par contre **certaines fonctions d'ATM sont très pratiques et ont permis d'avoir une infrastructure souple, facilement segmentable, pour véhiculer des flux divers qui ne se perturbent pas**; problématique des opérateurs de réseaux longues distance partagés entre plusieurs utilisateurs indépendants. Ainsi ces fonctions basiques d'ATM qui a été conçu initialement pour ces opérateurs de WAN, sont très utilisées sur ce réseau métropolitain. Ainsi :

. La possibilité d'avoir plusieurs VPs entre chaque site permet d'attribuer facilement des bandes passantes réservées de différents débits entre les sites, mais aussi pour certaines applications ou certains sous-réseaux IP, c'est à dire certains groupes d'utilisateurs. Il suffit dans ce cas de configurer correctement le routage ATM avec les numéros de VP associés. Cela permet aussi de séparer le trafic expérimental et de production sans que l'un perturbe l'autre.

. La facilité sur une même infrastructure ATM de créer autant de réseaux logiques IP (Classical IP) ou Ethernet (LANE) que l'on veut permet de créer des réseaux virtuels

entre communautés ou applications très facilement. Ainsi une station peut basculer d'un réseau logique à l'autre simplement avec une ou deux commandes, voire appartenir à plusieurs réseaux logiques. Dans la même idée, l'utilisation de PVC a permis de créer un réseau natif IPv6 entre certains équipements, c'est à dire un réseau avec un autre protocole que IPv4 (version IP actuelle). Cela peut-être nécessaire en particulier pour la transition IPv4 vers IPv6.

. Le protocole de routage ATM PNNI sur ce réseau maillé permet d'avoir une utilisation optimale, performante et robuste des liens, en adaptant le routage à la charge et à la disponibilité des liaisons de manière tout à fait transparente.

. L'utilisation de LANE avec une couverture régionale ou nationale, crée l'équivalent d'un réseau Ethernet, utilisable par les applications multicast. On dispose ainsi « en natif » d'une infrastructure de diffusion sans besoin de tunnels souvent difficiles à mettre en place comme sur les routeurs « classiques » Internet. Avec cette architecture de niveau 2, il est facile de monter des vidéoconférences pour une utilisation expérimentale ou pour un petit nombre de stations. C'est aussi une alternative aux tunnels IP pour créer l'équivalent d'un MBONE régional pour interconnecter les routeurs des sites et transporter des flux multicast entre ces sites.

. ATM peut permettre d'interconnecter des PABX très facilement (sous réserve des problèmes de synchronisation sur AAL5), sans que le trafic de données ne perturbe celui de la voix.

Ces utilisations très diverses du réseau auraient été très difficile à réaliser avec un réseau constitué de routeurs IP et de liaisons spécialisées classiques.

Par contre, **ATM est très loin d'être parfait** :

. Le coût des équipements (5 KF minimum pour une carte de station) a fait que l'utilisation d'ATM s'est limitée au coeur du réseau et ne s'est pas étendue à l'intérieur des sites, ce qui est un très bon choix économique.

. La technologie ATM par sa complexité nous a posé de grosses difficultés pour résoudre les problèmes de mise en place du réseau et est encore actuellement problématique lorsque quelque chose « ne marche pas ». Il est alors très difficile d'identifier où se situe le problème : lien physique, VP, routage IP, routage ATM, signalisation, shaping, serveur ARP, serveur LES-LECS, ... ? Il est nécessaire d'avoir une bonne méthodologie et un réseau d'experts.

. Une lacune flagrante d'un réseau élargi ATM est l'unicité des serveurs ARP d'un côté pour Classical IP et LECS, LES, ... de l'autre pour LANE. Cela implique que le réseau dépend de la bonne marche et de l'accessibilité d'un seul équipement. On est ainsi conduit à réduire la taille des réseaux IP et à créer plusieurs réseaux IP, forçant le passage à travers un ou plusieurs routeurs pour communiquer, ce qui réduit les performances par rapport à un chemin direct ATM. Ainsi, il aurait été plus performant d'avoir un seul réseau IP national pour avoir toujours des connexions directes ATM, sans passage par un routeur. Cela n'est pas raisonnable. On ferait dépendre le trafic interne C3I2 par exemple d'un seul serveur, certainement externe. La solution a ce problème peut être MPOA ou MPLS mais cela n'est pas encore disponible avec des équipements hétérogènes, comme c'est notre cas sur C3I2.

. La solution mise en place avec un maillage complet des n sites a l'inconvénient de ne pas supporter une mise à l'échelle puisqu'en moyenne le débit effectif vers chaque est de $134/(n-1)$ Mbps.

En conclusion on pourrait dire que la technologie ATM est une bonne solution de réseau métropolitain ou national quand la bande passante n'est pas illimitée, quand on

veut créer des réseaux indépendants (pour garantir des bandes passantes, créer des réseaux virtuels, mixer plusieurs protocoles, le téléphone et les données, ...), mais que sa complexité fait qu'elle doit être utilisée de la manière la plus basique possible et qu'il est inutile d'offrir un service ATM aux stations des utilisateurs.

6. Expériences pour la mise en place d'un réseau expérimental et d'un réseau de production multi-partenaires

De ce projet, on peut aussi tirer un certain nombre d'enseignements sur la manière de mettre en place un réseau expérimental ainsi que son utilité, et sur l'administration nécessaire pour un réseau de production de type privé avec plusieurs partenaires (notion de GFU, groupe fermé d'utilisateurs).

Coordination et organisation

Dans la conduite d'un projet de réseau expérimental limité à un an, il est d'abord impératif de **tenir compte du calendrier**. Pour différentes raisons (indisponibilité des matériels avec les fonctions nécessaires et inversement pression de France Télécom pour ouvrir le service rapidement), nous avons été obligé d'ouvrir ce réseau en décembre. Ainsi, le créneau est très défavorable, en particulier la coupure des vacances de l'été près de la fin du contrat a été très pénalisante pour la dynamique générale.

La mise en place de ce réseau a demandé une coordination humaine importante. Au niveau ingénieurs (les personnes qui administrent les équipements de communication), il est impératif que les équipes qui dépendent d'organismes différents apprennent à travailler ensemble, à échanger des informations, à se faire confiance et soient volontaires pour jouer le jeu des expérimentations. C'est d'autant plus nécessaire que ces administrateurs de réseau participe à ce travail en plus de leur charge de gestion quotidienne de leur réseau de production. Une telle synergie s'est réalisée mais il a nécessité une période de rodage. Pour les expérimentations entre équipes de recherche différentes, là aussi il faut du temps pour que des projets concrets prennent forme, pour que techniquement la connexion des équipements entre les 2 laboratoires soient réalisés et que les codes des applications soient prêts. Il est évident que les laboratoires manquent de bras pour expérimenter le fruit de leur recherche sur un « vrai réseau », car le chemin est long entre la maquette qui tourne sur un poste de travail et un outil utilisable à travers un réseau multi-sites.

Pour ce travail de coordination, de mise en contact des équipes de recherche, de mise en place des expérimentations et de résolution des difficultés techniques **il aurait été nécessaire d'avoir du personnel à plein temps affecté à ce projet**. Ainsi la résolution des problèmes de mise en place décrits au chapitre 3 ont mis en moyenne plusieurs semaines à être résolus car il n'y avait personne pour « s'y plonger » deux jours sans interruption.

Avec le basculement d'un trafic de production sur l'infrastructure cette nécessité s'amplifie. Dans l'administration d'un réseau métropolitain de production de type privé (l'opérateur externe n'assurant qu'un service de base), où ce sont les équipes des sites qui administrent les équipements, il ne faut pas sous estimer l'équipe d'exploitation nécessaire. Il est aussi nécessaire de bien définir le domaine d'intervention de chacun sur les différents équipements et essayant de bien doser la répartition de l'administration

entre un centralisme qui peut répondre garant de la cohérence, qui peut agir rapidement et efficacement face aux problèmes et une partie décentralisée pour laisser une certaine liberté et souplesse à chaque site.

Attitude scientifique

Lors des nombreux problèmes de mise en place rencontrés beaucoup de détracteurs de la technologie ATM ou simplement du projet ont crié immédiatement que « le réseau ne marchait pas », mettant en cause généralement le service ATM de base alors que le problème venait de la périphérie (carte ATM, OS, ...). La leçon est qu'il faut toujours adopter une attitude scientifique devant un problème de réseau jamais simple, où les éléments impliqués dans le processus sont très nombreux et prendre du temps pour diagnostiquer précisément la cause de l'anomalie et la solutionner.

ATM dans un réseau métropolitain ou régional

La mise en place de cette infrastructure et les expérimentations permettent maintenant d'avoir une idée précise de ce qu'il est possible et à contrario très difficile de faire avec un réseau ATM (cf chapitre précédent). Avant la mise en place de C3I2, la connaissance n'était que théorique, maintenant chaque ingénieur a mis en route et pratiqué cette technologie. Ainsi le design de réseaux métropolitains ou régionaux comme ARAMIS 2 sera plus facile à réaliser avec cette expérience acquise.

Outil de formation et de recherche

Mettre en place, utiliser un réseau expérimental permet d'acquérir réellement un savoir faire sur les nouvelles technologies de réseau mais aussi sur les applications du futur, apport difficilement quantifiable mais certainement très important pour les personnels des organismes qui participent à ces expérimentations.

Dans la communauté des chercheurs, cet outil donne de nouvelles idées d'étude et de collaboration à des équipes de recherche et permet une expérimentation en vraie grandeur sur un réseau.

Ces apports sont d'autant plus importants que le service réseau demandé à l'opérateur externe est minimal, comme des VPs ATM ou de la fibre optique nue. La période de mise en route est alors plus longue, l'exploitation est plus lourde, mais les protagonistes sur les sites sont vraiment confrontés aux problèmes de base et ont beaucoup de liberté de tests sur les protocoles, l'architecture de réseau, ... Ce service minimal permet aussi d'appréhender et de mieux expliciter les besoins télécom difficiles à évaluer dans notre milieu, même à court terme. Avec d'autres services, plus clé en main, on n'aurait pas par exemple pu se rendre compte que l'on avait surtout besoin de bande passante garantie, pour IP et pas d'ATM natif.

Le fait de disposer d'un réseau expérimental (ou d'une bande passante réservée pour des expérimentations sur un réseau de production), permet d'utiliser des nouvelles applications sans risque de perturber le réseau de production, élément maintenant vital. Il aurait été impossible par exemple d'expérimenter certaines applications de vidéoconférence à travers ARAMIS avec des accès de site trop chargés, et pire d'utiliser des versions expérimentales de logiciel comme c'est le cas sur certains équipements de C3I2.

Pour être un peu en avance sur la technologie, être précurseur sur les usages en matière de réseau, et faire partager ces expériences aux industriels et au grand public, la communauté de la recherche a besoin de projets de ce type.