

Ambassade de France aux Etats-Unis
MISSION POUR LA SCIENCE ET LA TECHNOLOGIE

LES RESEAUX A TRES HAUTS DEBITS ET LE MIDDLEWARE POUR LES GRILLES DE CALCUL		
FEVRIER 2002		

RAPPORT DE MISSION A WASHINGTON DE

- MICHEL COSNARD (INRIA SOPHIA ANTIPOLIS), RESPONSABLE
- VICTOR ALESSANDRINI (CNRS-IDRIS)
- FREDERIC DESPREZ (INRIA, ENS LYON)
- PASCALE PRIMET (ECL/UCB LYON)
- THIERRY PRIOL (IRISA/INRIA RENNES)
- CATHERINE ROUCAIROL (MINISTERE DE LA RECHERCHE)
- DANY VANDROMME (RENATER)

SOUS LA RESPONSABILITE DE

LAURENT DE MERCEY
ATTACHE POUR LA SCIENCE ET LA TECHNOLOGIE
WASHINGTON

Rapport de mission à Washington

sur les réseaux à très haut débit et le middleware

pour les grilles de calcul

A l'initiative de la Mission pour la Science et la Technologie de l'Ambassade de France aux Etats-Unis, une mission d'expertise a été conduite, les 30 et 31 octobre 2001 à Washington, pour mieux connaître les initiatives américaines dans le domaine des réseaux à très haut débit et du middleware (ou logiciels de base utiles à différentes applications) pour les grilles de calcul. La délégation française était composée de :

- Michel Cosnard (INRIA Sophia Antipolis), responsable
- Victor Alessandrini (CNRS-IDRIS)
- Frédéric Desprez (INRIA, ENS Lyon)
- Pascale Primet (ECL/UCB Lyon)
- Thierry Priol (IRISA/INRIA Rennes)
- Catherine Roucairol (Ministère de la Recherche)
- Dany Vandromme (Renater)

La délégation a rendu visite à Internet 2, à la National Science Foundation (NSF) et au Department of Energy (DoE). Lors de chacune de ces visites, les interlocuteurs américains nous ont présenté leurs projets, puis les participants français ont fait des présentations de leur côté qui ont servi de base aux échanges de vue et aux discussions. Nous souhaitons remercier l'ensemble des participants américains pour leur disponibilité, leur accueil et la qualité de leurs présentations, ainsi que la Mission scientifique pour la parfaite organisation matérielle.

Ce rapport est constitué de deux grandes parties :

- les réseaux à très haut débit,
- le middleware pour les grilles de calcul

Dans chacune de ces parties, nous faisons le point sur les projets en cours aux USA qui nous ont été présentés. Puis nous les plaçons en perspectives avec des projets en France ou en Europe et nous terminons par des recommandations.

Table des Matières

I Les réseaux à très haut débit

I.1 Aux USA

I.1.1 vBNS - NSF

I.1.2 ABILENE

I.1.3 ESNET

I.1.4 Nœuds d'échange : STARTAP et STARLIGHT

I.2 Réseaux académiques de recherche en France et Europe, réseaux pour les grilles

I.3 Recommandations

II Les initiatives middleware pour les grilles de calcul

II.1 Aux USA

II.1.1 Internet 2

II.1.2 NSF

II.1.3 DoE

II.2 Comparaison avec les travaux français et européens.

II.2.1 Action concertée incitative GRID : Globalisation des ressources informatiques et des données

II.2.2 Les projets RNRT et RNTL

II.2.3 Les projets européens

II.3 Recommandations

II.3.1 Middleware

II.3.2 Partenariats stratégiques

ANNEXE A Abréviations

ANNEXE B Etapes du réseau vBNS (et vBNS+)

ANNEXE C Cartes du réseau vBNS/vBNS+

ANNEXE D Cartes du réseau Abilene

ANNEXE E Cartes du réseau ESNET

ANNEXE F Projets financés par le DoE

I Les réseaux à très haut débit

I.1 Aux USA

Les réseaux à très haut débit ont été les premiers à profiter de la volonté politique des gouvernements américains pour développer les Autoroutes de l'Information. Il faut préciser également que lors du lancement de ces projets d'Inforoutes, il n'était absolument pas clair que le support en serait un développement accéléré de l'Internet. La convergence entre Internet et Autoroutes de l'Information n'a été entérinée qu'après environ un an du premier exercice de l'administration Clinton. A partir de cette date, le gouvernement américain a commencé à prendre des initiatives d'importance pour développer les infrastructures et les usages des réseaux à très haut débit. Plusieurs initiatives complémentaires ont alors pris corps : Next Generation Internet (NGI) et Internet2

Sous le vocable NGI, se sont développées les actions des agences fédérales (thématiques) comme la NASA (NREN), le DoE (ESNET), le Département de la Défense (DREN), etc.

Sous l'appellation Internet2, il y eu d'abord le réseau vBNS, qui correspond à un retour de la NSF, après l'échec du transfert de NSFNET au secteur commercial, puis le réseau Abilene. Alors que vBNS s'est fait dans le cadre d'un partenariat entre la NSF et MCI-Worldcom, le réseau Abilene s'est monté dans un partenariat entre un consortium d'universités (UCAID : University Consortium for the Advanced Internet Development) et l'opérateur QWEST. Dans les deux cas, les équipementiers participent de façon significative (Nortel, Cisco, Alcatel, etc.).

En accompagnement du développement massif de l'Internet, l'administration américaine a été consciente de l'importance des investissements dans les technologies et les usages les plus avancés. Cette démarche politique a le double objectif de maintenir son dispositif pour l'éducation et la recherche à la pointe du progrès, tout en induisant une situation de quasi monopole pour l'industrie américaine des télécommunications (infrastructures et équipements). Cet accompagnement prend des configurations qui évoluent au cours du temps, mais qui maintiennent les USA à la pointe du progrès dans ces domaines.

Les trois visites qui ont été effectuées dans le cadre de cette mission ont permis de rencontrer les acteurs de Abilene (UCAID/Internet-2), vBNS (National Science Foundation) et d'ESNET (DoE).

I.1.1 vBNS - NSF

Après l'expérience malheureuse de l'abandon de NSFNET au secteur marchand, la NSF s'est focalisée sur un réseau d'interconnexion de quelques grands centres de calcul (San Diego, Cornell, Pittsburgh et Boulder). Ce réseau s'est rapidement ouvert à un certain nombre d'universités, qui avaient des besoins évidents de raccordement à ces centres de calcul. Le réseau vBNS (« Very Broadband Network Service ») est issu d'un partenariat équilibré (50/50) entre la NSF et MCI-Worldcom, pour fournir une infrastructure utilisant IP/ATM pour l'interconnexion. Ce réseau, est basé sur des liens à haut débit entre des PoP (155 → 622 Mbit/s). Des équipementiers interviennent également, mais plutôt sous forme de partenariats

directs avec les sites. L'annexe B, qui est extraite du site de présentation du réseau (www.vbns.net) retrace les principales étapes de l'évolution de cette infrastructure. Pour mettre en évidence le modèle de partenariat, entre la NSF et MCI, avant le terme de la première tranche de 5 années, la technologie validée par MCI dans le service vBNS, a été transformée en une offre commerciale (vBNS+), proposée à toutes les universités, mais également à tout client (entreprise, organisation etc..). Le partenariat avec la NSF a été reconduit jusque 2003, pour supporter l'accès à vBNS+ pour un certain nombre d'universités.

I.1.2 ABILENE

Le second réseau d'infrastructure, qui entre sous le chapeau du projet « INTERNET-2 » est le réseau Abilene. Il a été lancé à l'initiative d'un certain nombre d'universités, sous forme d'une « réaction » à l'initiative fédérale (NSF). Ces universités ont créé un consortium (UCAID : University Consortium for Advanced Internet Development), qui a profité d'un partenariat avec l'opérateur QWEST (un nouvel entrant du secteur des télécommunications, dont le cœur de métier initial était le déploiement de fibres optiques le long des lignes de chemins de fer). Ce projet a fait un choix de technologie radicalement différent de celui de la NSF, en raison principalement de la disponibilité de bandes passantes inégalées (= 2.5 Gbit/s). Ce réseau a ainsi marqué le lancement à grande échelle de IP/SONET, en s'appuyant sur un partenariat très fort avec les 3 principaux équipementiers présents aux US : CISCO, NORTEL et ALCATEL pour les technologies IP et optiques. Ces sociétés font par ailleurs partie du consortium UCAID (tout comme QWEST), avec un statut de membre industriel. L'annexe D montre la topologie physique du réseau, qui est naturellement très semblable à celle de vBNS.

Les caractéristiques principales du réseau Abilene, en dehors de sa technologie 100 % IP, sont de s'appuyer sur un certain nombre de Gigapops, situés sur tout le territoire. Charge aux universités de se raccorder au gigapop le plus proche.

Une autre caractéristique marquante du réseau est son évolutivité constante. Alors que dans sa configuration initiale, cette initiative était strictement universitaire, elle s'est ouverte progressivement aux partenariats industriels (avec les partenaires membres du consortium UCAID) comme ALCATEL qui a ouvert un laboratoire au Texas raccordé à Abilene, ou des sociétés comme FORD (automobile) ou encore Johnson & Johnson. La liste des partenaires est disponible sur le site web de UCAID (www.internet2.edu).

Autre signe d'ouverture, le projet QUILT (<http://www.thequilt.net/>) qui s'appuie sur les initiatives régionales (réseaux régionaux ou universitaires) pour drainer sur Abilene, les flux en provenance ou à destination des établissements d'éducation primaire ou secondaire. Cette initiative montre le virage franc qui est pris par Abilene pour s'ouvrir largement en dehors de la stricte population des universités ! Grâce à l'interconnexion des réseaux des participants et un forum d'échanges et de communication, les missions du Quilt sont bien sûr d'expérimenter et de promouvoir des services avancés aux communautés E&R sur tout les Etats-Unis, d'effectuer des opérations de métrologie pour améliorer la connexion jusqu'à l'utilisateur final mais surtout, point remarquable, de représenter l'ensemble des intérêts communs des participants face aux fournisseurs de services, Internet, etc.

L'animation qui est faite par UCAID est tout à fait remarquable. Au-delà d'une fonction de maîtrise d'ouvrage technique du réseau de backbone, UCAID joue un rôle particulièrement efficace en termes de promotion et de présentation, qui est possible par les ressources humaines mobilisées (plus de 50 personnes).

I.1.3 ESNET

Le troisième réseau dont il a été question lors de la visite au DoE (Department of Energy) est ESNET. A la différence des deux réseaux universitaires précédents, ce dernier est un réseau thématique, dédié exclusivement aux laboratoires travaillant dans le domaine des « Sciences de l'énergie », en fait sous la tutelle directe du département correspondant de l'administration américaine.

On retrouve sur le réseau ESNET, toutes les communautés liées aux communautés IN2P3 en France, mais également celles travaillant sur les problèmes de fusion/fission etc. Ce réseau est géré par le personnel de LBNL (Lawrence Berkeley National Laboratory) en délégation du DoE. Il fournit des capacités d'interconnexion à très haut débit à tous les sites de recherche du DoE, et assure également l'interconnexion avec les autres réseaux pour la recherche (plus de 100 réseaux interconnectés). L'objectif, tel que revendiqué par le DoE est de fournir à tous les scientifiques de ces disciplines, un accès unique à toutes les infrastructures de recherche, de calcul ou de traitement de l'information, indépendamment de la localisation géographique et temporelle des intéressés. Les informations spécifiques à ce réseau sont évidemment disponibles au public (<http://www.es.net>). Une représentation du réseau est fournie dans l'annexe E de ce rapport.

La technologie employée aujourd'hui encore, est très semblable à celle de vBNS : IP/ATM, ce qui implique une limitation stricte à 2.5 Gbit/s pour le transport et des interfaces de service IP limitées à 622 Mbit/s. Au cours des discussions avec les représentants du DoE, il a été difficile de saisir exactement quelles étaient les capacités de transmission disponibles à partir d'un nœud d'échange avec les autres réseaux de la recherche (RENATER au STAR-TAP à Chicago par exemple), mais il n'est pas évident qu'il soit possible d'assurer des circuits dédiés (pour l'expérimentation BABAR au SLAC à Stanford par exemple) de capacités supérieure à 34 Mbit/s. Ce point est encore en cours de discussion avec les responsables du réseau, après la fin de la mission.

I.1.4 Nœuds d'échange : STARTAP et STARLIGHT

Afin de permettre une bonne interconnexion entre les différents réseaux américains, la NSF a décidé de déployer un nœud d'échange, qui serait commun à toutes les infrastructures. Dans sa première version, ce nœud d'interconnexion (STARTAP : Science, Technology And Research Transit Access Point) a été installé dans un point de présence d'un opérateur américain (Ameritech) à Chicago. Il est opéré par le personnel du laboratoire EVL (Electronic Visualization Laboratory) de Tom Defanti, à l'Université de l'Illinois à Chicago, en délégation de la NSF. La technologie employée est naturellement analogue à celle de vBNS, c'est-à-dire basée sur un transport de niveau 2 en ATM. Néanmoins, les différents réseaux américains s'y interconnectent via un routeur, afin de pouvoir gérer de façon plus efficace les différentes annonces de routes.

Pour répondre aux pressions internationales, qui se sont focalisées dans un groupe de travail mis en place par le G7, le GIBN (Global Interconnection of Broadband Networks), le gouvernement américain a décidé d'ouvrir le STARTAP aux réseaux recherche des autres pays (sous réserve de prendre en charge les coûts des liaisons pour aller jusque Chicago). Le nombre de réseaux étrangers présent au STARTAP est tout à fait remarquable. On y trouve, entre autres des réseaux d'Asie-Pacifique (Corée, Singapour, Japon, Australie, Taiwan),

d'Amérique Latine (Mexique, Brésil, etc) et d'Europe. Cinq réseaux européens sont aujourd'hui raccordés physiquement au STARTAP : NORDUNET (pays scandinaves et Islande), SURFNET (Hollande), RENATER (France), UICC (Israël) et CERN. Ces cinq pays bénéficient par ailleurs d'une prise en charge partielle de leur coût d'accès à Chicago, via un projet NSF appelé EUROLINK.

Néanmoins, en raison de limites technologiques évidentes, et du manque de perspectives d'évolution de l'ATM, en termes de capacité, les Américains ont décidé de créer un second nœud d'échanges, utilisant une technologie 100% IP, à base de commutateur gigaEthernet. Ce nœud est hébergé dans les locaux de l'Université du Nord-Ouest (NWU : NorthWestern University) à Chicago, et a pour nom STARLIGHT. Alors que STARTAP est limité à des débits d'accès de 622 Mbit/s, STARLIGHT permet de s'interconnecter à des débits supérieurs au Gbit/s. Parmi les premiers réseaux présents au STARLIGHT, il faut mentionner SURFNET, qui s'y interconnecte avec Abilene. A terme, tous les services d'interconnexion de STARTAP seront transférés au STARLIGHT (à l'exclusion évidemment du service ATM !). C'est le cas plus précisément des réseaux thématiques comme ESNET ou NASA-REN et DREN. L'interconnexion de vBNS au STARLIGHT semble plus incertaine et restera sans doute du ressort de MCI ou de la NSF, si leur partenariat sur vBNS est prolongé au delà de 2003.

Du côté des travaux de recherches sur les services avancés, les axes étudiés sont : le multicast, la qualité de service, IPv6 et les mesures. Des plates-formes expérimentales sont construites au-dessus des réseaux de production pour étudier ces services. Ainsi le Mbone pour le multicast, le Qbone pour la qualité de service, le 6Bone pour Ipv6 sont déployés. En ce qui concerne les travaux sur les modèles de Qualité de Service dans IP, il est à noter un arrêt du Premium Service sur Abilene. L'approche DiffServ classique avec la fourniture d'un Premium Service plus rapide et plus sûr que le service IP Best effort classique est remise en question. Ce rapport préconise le développement de services « égaux mais différents » et du service scavenger (beyond best effort) QBSS. L'étude des problèmes posés par le déploiement du Premium Service ainsi qu'une analyse approfondie des besoins des applications en terme de qualité de service sont nécessaires et seront conduites par le groupe Qbone de l'Internet2.

Par ailleurs, au début 2001, une initiative originale a été lancée dans le cadre d'Internet2, la "End-to-end Performance Initiative" afin de permettre une étude approfondie du problème relatif au fossé entre les performances offertes par les réseaux d'opérateur et les performances obtenues aux extrémités. Cette initiative s'appuie sur les partenaires d'Abilene et se concrétise en particulier autour du projet WEB100. Cette initiative est basée sur une infrastructure de mesure distribuée afin de permettre une compréhension rapide et effective d'un problème de limitation de performance de bout en bout. Des spécialistes de l'analyse de performances ont été engagés et ils initient des tests. Un autre objectif est la dissémination des meilleures pratiques d'ajustement des paramètres des protocoles de transport et de configuration des équipements d'extrémité.

I.2 Réseaux académiques de recherche en France et Europe, réseaux pour les grilles

Du point de vue des infrastructures de production, GEANT (ossature européenne) et RENATER (ossature française) offrent des performances équivalentes (voire supérieures) et

s'appuient sur les mêmes technologies que les réseaux vBNS, Abilene ou EsNET : IP sur SDH pour GEANT et pour RENATER 3. Alors que RENATER 2 et TEN155 étaient aussi en IP sur ATM. On remarque que le passage en IP sur SONET (ie SDH) est tout de même antérieur aux USA. En effet GEANT est mis en service au 1/12/2001 et RENATER 3 en 2002, alors que Abilene est déployé depuis 1998.

En terme de couverture géographique et thématique, la comparaison est très difficile à faire. Les différents réseaux américains ne couvrent que des sous-ensembles de populations de chercheurs, bien souvent en duplication. Si on regarde l'implantation des nœuds de vBNS et des giga pops d'Abilene, on réalise à quel point les éléments du réseau se superposent.

A l'opposé, en Europe nous avons réussi à fédérer autour d'une même infrastructure 31 réseaux pour l'éducation et la recherche. Cela permet un accès unique à toutes les disciplines en une fois, ce qui n'existe pas aux US. Par ailleurs, nous avons (encore incomplètement) réussi à mutualiser une bonne partie des infrastructures intercontinentales pour l'interconnexion de GEANT avec le reste du monde. Alors que jusqu'à présent, les pays européens finançaient eux-mêmes leurs infrastructures transatlantiques (pour atteindre New York ou Chicago), on assiste maintenant à un retournement de situation. C'est maintenant UCAID, qui envisage de déployer des liens intercontinentaux pour raccorder Abilene à GEANT en Europe. Avant la fin du premier semestre 2002, la capacité d'interconnexion pour le trafic recherche entre l'Europe et les US sera de 10 Gbit/s, et financée symétriquement par les deux continents !

Ce modèle est par ailleurs en cours de développement également entre l'Europe et l'Asie et l'Amérique latine.

Concernant les services avancés, multicast, qualité de service, Ipv6, l'état des développements en Europe et en France sera bientôt voisin. Le service Multicast est fourni sur GEANT et sur RENATER. Les services Premium sont en cours de déploiement sur GEANT. L'Europe s'est sensibilisée au développement de Ipv6 bien avant les Américains et les déploiements sont bien avancés chez nous. En particulier le projet 6NET verra la mise en place au niveau européen, d'un réseau de capacité 2.5 Gbit/s, dédié dans un premier temps au protocole IPv6, avant d'être complètement intégré dans GEANT. Il n'existe pas d'initiative de cette ampleur sur aucun des réseaux américains. Pour la métrologie, la France n'est pas encore dotée d'infrastructures de monitoring de bout en bout en raison de l'architecture à plusieurs niveaux (réseaux nationaux, régionaux, métropolitains, de campus etc) qui sont sous des responsabilités différentes. Néanmoins, un effort soutenu est fait pour intégrer dans tous les niveaux de réseau, les notions de métrologie fine et de qualité de service de bout en bout.

La France s'est doté d'un dispositif de plates-formes expérimentales mis en place et subventionné par le biais du RNRT (Réseau national de Recherche en télécommunications) :

- @IRSbone : plate-forme d'expérimentation de protocoles Internet de nouvelle génération (Ipv6) ;
- AMARRAGE : plate-forme pour réseaux actifs (utilisée pour le déploiement et la validation de nouveaux applicatifs et protocoles réseaux) ;
- PAROL : plate-forme répartie à objets selon principe du logiciel libre (propose une base de code initiale pour le développement) ;
- PCNS : plate-forme de communication par satellite du CNES (expérimentation et évaluation de l'apport potentiel de satellites géo-stationnaires) ;

- RENATER : plate-forme utilisant l'ossature des nœuds de présence en régions de RENATER et reposant sur le service VP/VPN Virtuel Private Network de RENATER.
- STENTOR : plate-forme fournissant un support satellitaire pour de nouveaux protocoles de transmission et de nouveaux services dans le domaine des télécommunications spatiales ;
- VTHD (Vraiment très haut-débit) en technologie IP/WDM, mise en œuvre et opérée par France Télécom, réseau à très haut débit (longueurs d'onde sur fibre optique 2,5 Gigabit/s en débit, Giga-routeurs) ; VTHD a été et est utilisée pour des projets de type « Calcul sur les grilles ».

Ces plates-formes sont un ensemble de logiciel et/ou matériel susceptible de fournir de services de télécommunications innovants dans le cadre de la recherche. Elles concernent soit de l'expérimentation (projet de recherche et développement pour développer un nouveau service de télécommunication, mis à disposition d'utilisateurs), soit de l'intégration (création de services innovants par intégration de composants ou systèmes). RENATER a pour mission de fédérer ce dispositif qui permet à de nombreuses équipes d'expérimenter leurs travaux. Six responsables de plates-formes participent actuellement au groupe de pilotage mis en place par le RNRT en liaison avec le RNTL ainsi que des représentants des Ministères.

Les propositions de plates-formes foisonnent donc en France (deux nouvelles viennent d'être labélisées par le RNRT pour l'expérimentation UMTS d'une part, et les validations multi-protocoles et multi-services d'autre part, une dizaine sont en attente au RNTL). La situation ne nous a pas semblé être la même aux Etats-Unis ; en ce qui concerne l'expérimentation pour des grilles, soit un réseau dédié est mis en place (par exemple, TERA facilities project), soit un middleware est à l'étude pour être supporté par un réseau (Internet2). Par contre, le projet DTF devrait offrir des débits de 40Gb/s avec des capacités de calculs en bordures capables de remplir les liens. Des expérimentations comparables à celles menées aux Etats-Unis (multicast, QoS ou Ipv6) ont été conduites en France ou sont en cours.

Les travaux menés par le groupe réseau dans le cadre du projet européen DataGRID, décrit plus loin, relèvent de la même problématique que ceux qui nous ont été présentés à Internet2, à la NSF et au DOE. En particulier :

- Déploiement d'une infrastructure de monitoring et de mesure efficace.
- QoS : premium service remis en question/ développement des services « égaux mais différents » et du service scavenger (beyond best effort) et étude approfondie des besoins des applications.
- Transfert haute performance : remise en question du protocole TCP et étude pour transport très volumineux sur lien très haut débit
- Travaux sur GridFTP (pour end to end performance) avec Netlogger.
- Développement de protocole de Multicast fiable pour la Grille.

Les travaux dans le cadre de E-TOILE/ ACI-GRID sont en phase avec les travaux qui sont menés/financés aux USA. Nous notons cependant que la technologie des « réseaux actifs » n'est pas étudiée dans le cadre du Grid Computing aux USA. C'est un intérêt qui demeure actuellement purement français, bien qu'un projet européen sur le sujet soit en préparation.

En conclusion, au niveau de la recherche en Réseau, les problématiques étudiées et l'avancement des travaux ne sont pas toujours équivalents entre l'Europe et les Etats-Unis. Nous avons en France un réseau expérimental à très haut débit depuis 2 ans. Les problèmes d'ingénierie et de difficulté d'obtenir des performances de bout en bout ont été bien

identifiées. Il n'existe pas, à notre connaissance, d'initiative publique comparable à la « end to end performance initiative » pour traiter le problème du décalage entre le débit offert par le réseau et le débit utile observé au niveau des applications, en particulier en France.

I.3 Recommandations :

- Maintenir une veille technologique importante pour suivre notamment le projet DTF.
- Participer à des travaux internationaux (projets et normalisation IETF et GGF) pour ne pas être ignorés et pour ne pas refaire la même chose.
- Participer aux travaux européens (TERENA, SEQUIN)
- Poursuivre les déploiements de plates-formes pour expérimenter en vraie grandeur les technologies et les outils. C'est le seul moyen permettant d'étudier le passage à l'échelle des solutions proposées.
- S'appuyer sur des équipes d'ingénieurs solides. Ces travaux de recherches nécessitent des infrastructures énormes et la mise en oeuvre est souvent ingrate pour les chercheurs.

II Les initiatives middleware pour les grilles de calcul

Il est difficile de donner une stricte définition des grilles de calcul. Il s'agit de systèmes répartis (architecture distribuée) dont les applications sont essentiellement le calcul scientifique. Cependant, le spectre des applications des grilles s'élargissant (on a même parlé de vidéo-conférence lors de la visite à la NSF), il devient de plus en plus difficile d'identifier précisément ce que sous-entend le concept de grille. Pour les auteurs de ce rapport, le concept de grille est identifié à des systèmes répartis haute-performance dont certains des constituants sont des calculateurs parallèles. L'aspect le plus important dans ce type de systèmes est la conception de *systèmes d'exploitation répartis* permettant l'accès transparent aux ressources de calcul et aux réseaux et la gestion de celles-ci. Ce type d'infrastructure étant par nature très hétérogène, la conception d'un système d'exploitation n'est possible que par la mise en œuvre de services au dessus des systèmes d'exploitation existants (Unix, Windows, ...) sur chaque nœud. Ces services sont utilisés par les applications pour faciliter leur déploiement. On parle alors de « middleware ». C'est pour le moment l'approche la plus raisonnable pour concevoir des grilles de calcul réellement utilisables. Il ne faut cependant pas exclure que d'ici quelques années les systèmes d'exploitation classiques disposent de services en leurs seins permettant de concevoir des grilles de calcul plus facilement. Mais ce n'est que de la spéculation pour le moment sachant qu'il n'existe pas encore de modèle économique sous-jacent à la grille.

Le concept de grille de calcul, au-delà du simple marketing et de son effet « pompe à finance » aux USA et maintenant en Europe, est fondamental pour développer la culture « réseau » parmi les concepteurs et les utilisateurs d'applications. Il s'agit de les familiariser aux technologies liées à la sécurité, à l'utilisation efficace des réseaux ou bien à l'utilisation des « middleware ».

II.1 Aux USA

II.1.1 Internet 2

Le consortium Internet2 a lancé une initiative « Internet2 middleware initiative – I2-MI»¹, dotée de 12 millions de dollars et dirigée par Ken Klingenstein, pour le développement de projets dans le domaine du « middleware ». L'objectif de cette initiative est d'encourager le déploiement de services au sein d'une infrastructure de type « middleware » permettant le développement et l'utilisation d'applications à grande échelle exploitant les capacités des réseaux de dernière génération. Un accent particulier est mis sur l'interopérabilité de ces différents services entre plusieurs sites. Le rôle du consortium dans ce domaine est aussi d'encourager l'utilisation de standards, de stratégies d'implémentation et de disséminer les informations pour éviter la duplication des efforts. Il existe d'ores et déjà plusieurs activités au sein de cette initiative :

- *Middleware Architecture Committee for Education (MACE)* : un groupe d'experts dans le domaine de l'informatique et des télécommunications ayant pour objectif d'établir des directions techniques que devront suivre les projets ayant le soutien de I2-MI. Ce comité pilote essentiellement trois groupes de travail sur des domaines qui ont été identifiés par les experts comme critiques :
 - o Répertoire : il s'agit de construire des répertoires de données, un type de base de données optimisé pour des accès en lecture, nécessaires au fonctionnement de certains services middleware. Un accent particulier est mis sur la sécurité. Ce groupe de travail gère plusieurs projets : MACE-dir, DoDHE, eduPerson et LDAP Recipe
 - o PKI (Public Key Infrastructure) : il s'agit de faire avancer des travaux autour de l'autorisation et l'authentification dans des réseaux à grande échelle avec comme sujets prioritaires : interopérabilité, modèle de délégation, passage à l'échelle, ... Internet2 participe à deux activités dans ce domaine : HEPKI et I2 PKI Labs.
 - o Accès sécurisé au Web : l'objectif est de permettre l'accès à des pages web sécurisés sur des sites utilisant des mécanismes d'authentification variés sans obliger à adopter le même mécanisme. Les projets Shibboleth, en collaboration avec IBM, et le projet WebISO adressent en particulier cet aspect.
- Organisation d'ateliers techniques à destination des ingénieurs en charge des infrastructures informatiques dans les universités. Il s'agit essentiellement d'encourager l'adoption de middleware et de disséminer les expériences acquises entre plusieurs universités.
- Participation à des projets sur les grilles de calcul comme le projet Beta Grid financé par la NSF Middleware Initiative – NMI. Il s'agit d'un grand projet sur le développement d'une grille de calcul au sein duquel Internet2 a pour rôle de fournir les technologies de répertoire, de sécurité et de nommage. Internet2 fournit également du support aux projets de grille en mettant à disposition du matériel réseau, en effectuant des présentations techniques (comme par exemple le projet NEES Grid)

L'approche d'Internet2 est très pragmatique. Plutôt que de tenter d'imposer une technologie particulière pour le déploiement d'une grille de calcul, le consortium a sélectionné trois domaines essentiels qui ne concernent pas uniquement la conception de grille de calcul : répertoire, sécurité et accès sécurisé. Le rôle d'Internet2 est d'encourager l'expérience de systèmes répartis à grande échelle et non pas uniquement l'utilisation de grilles de calcul pour des applications du calcul scientifique. Il s'agit d'une approche très pragmatique : déployer

¹ <http://middleware.internet2.edu/>

tout d'abord des mécanismes de base (répertoire, protocoles de sécurité, ...) puis une fois ces mécanismes déployés et les utilisateurs familiarisés avec ceux-ci, encourager leur utilisation dans le cadre d'applications (grille de calcul, ...). Il ne s'agit donc pas d'encourager le déploiement d'une infrastructure lourde, de type Globus, qui serait dans certains cas inutile pour les applications visées.

II.1.2 NSF

Les initiatives de la NSF concernant le middleware sont initiées par le Directorate for Computer and Information Science and Engineering (CISE) à travers principalement deux divisions : l'ACIR (Advanced Computational Infrastructure and Research) et l'ANIR (Advanced Networking Infrastructure and Research). Une des missions de la CISE consiste également à fournir l'accès à la puissance de calcul en donnant l'accès à des supercalculateurs connectés via de puissants réseaux à travers les Etats-Unis.

Dans l'ACIR, on peut remarquer tout particulièrement le projet PACI (Partnerships for Advanced Computational Infrastructure) dont le but est de fournir l'accès à des réseaux de recherche connectant de nombreux supercalculateurs. Néanmoins, ce programme finance également le développement de logiciels permettant de faciliter l'accès à cette grille et le portage d'applications adaptées à ce type de plate-forme. Par ailleurs, un des buts du programme PACI consiste à mettre en place le Terascale Computing System (TeraGrid) qui permettra l'exécution d'applications de physique (GriPhyN) et de simulation de tremblements de terre (NEES) grâce à la connexion de plusieurs machines de plusieurs Térafllops (Caltech, Argonne, SDSC, NCSA) via un réseau rapide.

La NSF a lancé en août 2001 le projet DTF Distributed Terascale Facility (53 Millions de \$) qui consiste à déployer des capacités distribuées à l'échelle tera : la TERAGRID. DTF se veut être l'infrastructure la plus large jamais été déployée pour la recherche scientifique avec plus de 13.6 Terafllops (trillions d'opérations par seconde) de puissance de calcul et de capacités pour gérer et stocker plus de 450 Terabytes (trillions d'octets) de données. Le projet DTF d'une durée de trois ans, consiste premièrement en des serveurs en cluster basés sur la famille des processeurs Itanium™ de Intel® interconnectés avec le réseau Myrinet de Myricom. Les clusters seront des nœuds distribués reliés via un réseau optique dédié à 40 Gigabits par seconde et qui sera plus tard upgradé à 50-80 Gigabits par seconde. Ce réseau, mis en oeuvre par Qwest sera connecté à Abilene, STAR TAP, et CalREN-2 (réseau très haut débit des universités californiennes). En Illinois, I-WIRE, le réseau optique fournira la capacité optique.

Dans l'ANIR, le projet NMI (Network Middleware Initiative) a retenu notre attention. Cette initiative a été initiée en février 2001 et a financé 9 projets autour du middleware. L'organisation du NMI est divisée en deux équipes : Internet2 et les GRID Centers (ISI, NCSA, UC, UCSD, UW). Le but ici est clairement affiché de concevoir, de développer et de déployer le middleware nécessaire aux applications portées sur la grille. La transparence d'utilisation est primordiale et les outils développés incluent naturellement les couches de gestion du réseau, des ressources et de la sécurité mais aussi les outils de collaboration entre les chercheurs (vidéo, enseignement à distance, etc.). Les premières applications cibles sont celles portées dans le cadre de la TeraGRID. Pour 2002, l'accent est mis sur le développement de couches logicielles pour la gestion de la sécurité (authentification, transferts de données) et sur l'analyse et le monitoring des performances de la grille.

Par ailleurs, la NSF a lancé le programme NGS (Next Generation Software). Il s'agit ici de développer des suites logicielles adaptées au portage d'applications sur la grille. Tous les niveaux de logiciels sont concernés, des couches basses proches de l'architecture jusqu'au compilateurs de langages de haut niveau en passant par la modélisation des applications et des architectures cibles. Le projet GrADS, présenté lors de notre visite par Jack Dongarra est financé par ce programme.

La liaison entre les différents programmes initiés par la NSF n'est pas claire. Même si on constate que la plate-forme elle-même est gérée par le programme PACI, les couches « basses » par le programme NMI et enfin les environnements logiciels plus complets par le programme NGS, de nombreux recouvrements existent.

II.1.2 DoE

Les activités autour du middleware et des grilles de calcul sont placées sous l'égide du « Office of Advanced Scientific Computing Research – ASCR » qui dépend du « Office of Science » du département de l'énergie. En 2001, l'ASCR, via le département MICS (Mathematical, Information and Computational Sciences) a lancé le programme SciDAC (Scientific Discovery through Advanced Computing) qui a pour objectif de financer des projets de recherche et de développement dans trois domaines essentiels : middleware haute-performance, réseaux haute-performance (protocoles de transport, outils pour l'analyse et la mesure de performance des réseaux) et applications coopératives.

A partir de la liste des projets financés par le DoE, il apparaît clairement un soutien fort et massif à la boîte à outils Globus (notamment dans le cadre de l'initiative SciDAC). Cette technologie est au cœur d'au moins sept projets au sein desquels elle sera utilisée comme plate-forme d'expérimentation. Le concurrent direct de Globus, Légion, n'apparaît plus et semble désormais n'être utilisé que dans des projets de nature plus académique. Il faut noter cependant que Légion fait l'objet d'un transfert technologique via la société AVAKI². Le DoE ne semble pas oublier d'encourager des expériences en utilisant les technologies du Web (XML, WebDAV, SOAP, ...) plus facile d'accès à des utilisateurs moins expérimentés mais ceci est réalisé au sein de projets plus petits (en terme de financement). Dans un objectif de renforcer l'utilisation des technologies développées par chaque projet, le DoE oblige les projets à lister leurs relations avec les autres projets financés dans le cadre de l'initiative SciDac et à décrire ce que font ces projets avec les technologies qu'ils développent. Enfin, il est important de souligner que le financement de projets semble très cohérent avec un bon équilibre entre le développement de middleware, l'expérimentation de protocoles de sécurité, l'utilisation efficace de réseaux et enfin le déploiement d'applications à grande échelle dans des domaines très variés.

II.2 Comparaison avec les travaux français et européens.

II.2.1 Action concertée incitative GRID : Globalisation des ressources informatiques et des données

² <http://www.avaki.com>

Dans un contexte international fortement compétitif où le développement de logiciels pour les plates-formes globales de calcul repose de plus en plus sur la capacité à intégrer des approches complémentaires mettant en jeu des disciplines et des compétences variées, le Ministère de la Recherche a lancé en 2001 une Action Concertée Incitative (ACI) afin d'amplifier les activités de recherche dans les sciences et technologies de l'information et de la communication sur les développements de logiciels permettant la conception et la mise en œuvre de ces infrastructures globales et de développer des actions pluridisciplinaires en vue de l'utilisation de ces équipements.

L'objectif de l'ACI est de dynamiser et rendre opérationnelle, à un moment critique, la contribution des équipes de recherche françaises dans le domaine, en soutenant les équipes déjà actives, en attirant de nouveaux acteurs, en favorisant les rencontres entre concepteurs de nouvelles solutions et utilisateurs. En s'appuyant sur des plates-formes de réseaux à haut débit fournies par Renater (le réseau national académique pour la recherche et l'enseignement) et le projet VTHD (plate-forme expérimentale à très haut débit) mis en œuvre dans le cadre du RNRT, et sur des centres de calcul nationaux et régionaux, sur des infrastructures régionales ou à l'échelle d'un campus, plusieurs axes de recherche pourront être développés et contribuer à la mise en œuvre de "grilles expérimentales" : déploiement d'outils logiciels (accès distribué, authentification, sécurité et gestion des échanges et des ressources, interfaces pour le "metacomputing" ou l'exploration de gisements d'informations, aide à la décision, etc.), systèmes et environnements pour le calcul distribué (logiciels de médiation, partage transparent de ressources,...) ou pour l'exploitation d'ensembles de données de très grande taille (dont bases de données textuelles, bases de données d'images, multimédia), modélisation, algorithmique, couplage de codes, visualisation, pré et post-traitements associant les approches théoriques des modèles dans les grands domaines scientifiques, les résultats expérimentaux et les approches de type simulation par calcul.

II.2.2 Les projets RNRT et RNTL

Le projet **E-Toile** est un projet de plate-forme du Réseau National de Technologies Logicielles (RNTL) d'une durée de deux ans (2002-2003) construit sur trois objectifs :

- opérer une grille de calcul expérimentale en France pour supporter des applications industrielles et scientifiques, parmi lesquelles les projets développés dans le cadre de l'ACI GRID,
- évaluer et mettre au point une base logicielle intégrant les composants logiciels développés dans les laboratoires de recherche français (middleware de grille),
- déployer et expérimenter une technique de réseau actif sur l'infrastructure VTHD au bénéfice de cette grille expérimentale.

Dans le cadre d'E-Toile, un support middleware sera fourni à la communauté des utilisateurs afin de leur faciliter l'accès aux ressources réparties. Une évaluation comparative des différentes solutions existantes est en cours afin de définir le meilleur support de base à déployer en tenant compte des objectifs expérimentaux de la plate-forme. En particulier, les services spécifiques, développés par les équipes de recherche française afin de répondre à des besoins particuliers d'applications et d'utilisateurs industriels, besoins non ou mal couverts (à court et moyen terme) par les middleware existants devront pouvoir s'intégrer aisément à la suite logicielle sélectionnée. La solution Globus a été initialement pressentie comme solution car c'est une plate-forme globale qui a le mérite d'exister et qui permet une initiation aux concepts de grille. Le choix définitif ne devrait être arrêté qu'après l'étude en cours.

II.2.3 Les projets européens

Les principaux projets européens dans ce domaine sont les projets DATAGRID, EUROGRID et DAMIEN.

Le projet **DataGrid** est un projet européen (5^{ème} PCRD) d'une durée de trois ans (2001-2003) visant à installer, développer et expérimenter une technologie de grille de calcul sur des infrastructures de réseau à très haut débit (Géant et liaison Gbits Europe-US) au bénéfice d'applications scientifiques. Les applications visées sont d'abord la physique des particules qui entend disposer de cette technologie pour le traitement massif des données du LHC, l'imagerie médicale, la génomique et les sciences de la terre (traitement de données satellite sur la couche d'ozone). La coordination des WorkPackages Plate-forme expérimentale, Réseau et Applications de Biologie est assurée par la France sous la responsabilité du CNRS.

II.3 Recommandations

II.3.1 Middleware

Au travers des projets financés par les différentes institutions (Internet2, NSF, DoE), les Etats-Unis sont la locomotive dans le domaine des grilles de calcul. La force de ces initiatives vient sans aucun doute de la capacité de ces institutions à faire travailler ensemble des chercheurs de plusieurs domaines (informaticiens, physiciens, chimistes, ...). Cette association a un effet mobilisateur important. Les problèmes à résoudre sont clairement identifiés et le déploiement d'applications à grande échelle largement motivé par l'étendue du territoire américain et la dispersion des grands centres nationaux de recherche. En France, il est encore très difficile d'avoir ce niveau d'association entre chercheurs de domaines différents, à l'exception notable du travail qui se fait depuis des années dans les centres nationaux où des ingénieurs informaticiens ont des collaborations très productives avec des spécialistes applicatifs. Cependant, dans le domaine de la recherche avancée et en caricaturant, le sentiment général des informaticiens est qu'un physicien ne voit pas souvent l'intérêt de s'associer à des chercheurs informaticiens car il peut lui-même faire le travail en prenant des stagiaires... D'autre part certains chercheurs se plaignent que les informaticiens ne s'intéressent que très peu à leurs applications. Cette situation devient de plus en plus difficile à maintenir compte tenu de l'évolution de l'informatique (le calcul parallèle est incontournable, l'utilisation de réseaux et donc de grilles de calcul) ; elle est de plus dommageable pour l'ensemble de la communauté scientifique. Les chercheurs physiciens, chimistes, ... auront du mal à se comparer aux chercheurs américains sur la validation de leurs modèles théoriques compte tenu qu'ils n'auront pas accès aux moyens de calcul nécessaires. D'autre part les chercheurs informaticiens sont coupés d'un bassin applicatif générateur d'idées et de concepts nouveaux. Il y a donc lieu à encourager la pluridisciplinarité dans les projets autour des grilles de calcul pour réduire cette séparation – très franco-française – entre les chercheurs des domaines d'application, les sites de production comme les centres de calcul – nationaux ou autres - et les structures de recherche en informatique (le couplage est bien plus fort aux USA, en Grande-Bretagne et en Allemagne). Une relation plus étroite entre les centres de super calcul et les centres de R&D en informatique, comme cela se fait dans les centres de puissance intermédiaire (Nancy, Rouen, Rennes, Toulouse) donnerait,

aux chercheurs informaticiens, un accès plus aisé aux codes grandeur nature et aux vraies applications scientifiques.

En ce qui concerne le déploiement d'infrastructure de grilles de calcul, la France a accumulé sans aucun doute un retard, pas seulement par rapport aux Etats-Unis mais aussi par rapport au Royaume-Uni (programme e-Science) et en Allemagne (au travers de leurs centres de calcul plus nombreux qu'en France, même si le nombre n'implique pas obligatoirement des résultats scientifiques de meilleure qualité). Remarquons cependant que les deux centres de calcul nationaux participent au projet européen Eurogrid. Elle a cependant des atouts avec une communauté scientifique assez forte dans les aspects systèmes (y compris middleware) et réseaux. Il manque cependant une réelle coordination entre les différentes initiatives (RNRT, RNTL, ACI, ...) qui financent des recherches dans le domaine des grilles de calcul. Le développement de liaisons entre ces programmes permettrait de présenter à la communauté scientifique internationale une vision cohérente de toutes les activités de recherche dans le domaine des grilles de calcul en France. Le démarrage récent d'actions incitatives (ACI GRID) va dans le bon sens mais il ne faudrait pas négliger son rôle de communication et de dissémination des travaux réalisés en France.

Une des questions qui se pose est quelle position doit-on adopter par rapport aux technologies de grille de calcul développées aux Etats-Unis (comme par exemple Globus) ? Nous n'avons pas beaucoup entendu parlé de Globus mais, à la lecture des documents fournis, il semblerait que Globus soit devenu un standard de fait. Il faut donc être vigilant et faire en sorte que les contributions françaises et européennes soient mises en valeur face à celles de Globus. Pour se positionner et susciter l'intérêt des différents acteurs Grid, il faut bien cerner et analyser les atouts et les limites des services offerts par Globus. Il ne faut cependant pas aller dans l'extrémisme. Il est important de comprendre que ces grilles sont avant tout des systèmes répartis. Globus est une boîte à outils essentiellement développée pour prendre en compte l'utilisation de supercalculateurs. Il ne faut donc pas hésiter à l'utiliser si les applications nécessitent de telles machines. Cependant, il ne faudrait pas décourager l'utilisation de systèmes répartis un peu plus modestes. Ces initiatives autour des grilles de calcul doivent être aussi l'occasion de promouvoir l'utilisation de ces systèmes afin de propager une culture « réseaux » dans la communauté scientifique. De nombreuses technologies (sous forme de standards) existent d'ores et déjà et méritent d'être utilisées en dehors des champs d'application pour lesquelles elles ont été développées. On peut citer par exemple : CORBA, Java (notamment les EJB), SOAP, XML-RPC, UDDI, WSDL, ... Le développement de projets autour des grilles de calcul doit être également un véhicule pour mieux faire connaître l'existence de ces technologies.

A plus long terme, il sera indispensable de rechercher l'interopérabilité entre toutes ces technologies. Ceci passe par la création d'un organisme de standardisation fort et crédible dans le domaine des grilles de calcul (ce que n'est pas encore le Global Grid Forum). Il peut paraître inopportun de mettre trop tôt l'accent sur ce point surtout à court terme car cela nous enverrait systématiquement vers l'interopérabilité avec Globus, faisant de celui-ci un standard de facto. L'implication plus forte des chercheurs français dans le domaine des structures de lobbying ou de standardisation doit être encouragée (et aussi reconnue par les instances d'évaluation !).

Le Grid Forum, à l'initiative des chercheurs américains, et maintenant le Global Grid Forum (incluant les Européens et le Japon), ont pour objectifs d'établir des standards et d'améliorer l'interopérabilité des infrastructures de grilles de calcul. Malheureusement, l'intégration des

Européens dans le Global Grid Forum, via l'initiative Egrid, a été prématurée (quelques mois après sa création) ce qui n'a pas permis aux européens de peser vraiment dans la direction de ce forum (il suffit de voir les agendas de ces réunions dans lesquelles interviennent majoritairement des chercheurs américains). Dans de nombreux domaines (réseaux, middleware, sécurité, modèle de programmation), des initiatives ont été prises pour faire avancer des standards. Il y a par exemple des initiatives auprès de l'IETF (AccessGrid) du W3C et de l'IRTF. En amont de ces initiatives, le DoE participe activement au financement de projets visant à concevoir des modèles de composants logiciels pour les grilles de calcul (CCA Forum). Il s'agit de spécifier un modèle permettant d'encapsuler les codes de calculs, notamment ceux développés dans les projets financés par le DoE, au sein de composants. L'idée est de pouvoir composer une application sur une grille de calcul à partir de composants disponibles sur plusieurs nœuds de la grille. Il s'agit ici d'un aspect stratégique qui vise à permettre la réutilisation de codes existants. Il y a fort à parier que ce modèle sera imposé par le DoE dans l'avenir. Bien entendu, il existe d'autres initiatives de standardisation des modèles de composant (OMG CCM, Sun EJB) mais ils ne sont pas a priori conçus pour des grilles de calcul. Pour motiver cette remarque, il suffit de lire le résumé du projet CCTTSS (Center for Component Technology for Terascale Simulation Software financé par la DoE dans le cadre du programme SciDAC) et notamment les relations avec les autres projets « applicatifs » de ce programme.

En conclusion, même si les laboratoires américains sont très moteurs sur le développement de logiciels de tous niveaux autour de la grille et la mise en place de plates-formes d'expérimentation de taille conséquentes, la recherche française a clairement un rôle à jouer avec des relations plus franches avec la communauté du système et des réseaux et sur des aspects plus fondamentaux (algorithmique, ordonnancement, placement des données, modélisation).

II.3.2 Partenariats stratégiques

Lors des discussions avec Internet 2, nous avons remarqué la politique des Partenariats Stratégiques avec des compagnies qui apportent 1 million des dollars au budget de l'institution. Certaines de ces compagnies sont des fournisseurs des biens et des services potentiels. La question de la gestion d'éventuels conflits d'intérêt se pose à l'observateur européen.

La première observation est que cette politique des partenariats stratégiques n'est pas limitée à Internet 2. Les centres de calcul de la NSF la pratiquent depuis longtemps. Les partenaires stratégiques émanent de l'industrie automobile, aéronautique, chimique, mais aussi des prestataires des services réseau et des constructeurs informatiques.

Une telle association serait impensable en Europe : imaginer IBM apporter 1 million de dollars à l'IDRIS ? Mais aux USA les appels d'offres pour les gros projets nationaux ne se passent pas de la même manière qu'en Europe. Lorsque la NSF lance un appel d'offres pour une « terascale facility », les centres de calcul s'associent à un constructeur et à un ensemble de fournisseurs pour répondre. Fin 1999, la première « terascale facility » a été gagnée par le couple Pittsburgh – Compaq face à San Diego – IBM et NCSA – IBM. Celle qui vient d'être décidée résulte du partenariat de San Diego, Caltech, Argonne et NCSA avec IBM.

Dans cette optique, les partenariats stratégiques prennent tout leur sens. Les choix stratégiques sont déterminés par l'optimisation d'un intérêt national beaucoup plus large que le projet

particulier. Cela implique bien évidemment que les solutions techniques choisies pour tel ou tel projet particulier ne sont pas forcément optimales. Il est important de bien comprendre ce point particulier pour garder une distance prudente vis-à-vis des choix stratégiques américains : la composante politique joue certainement, et ils l'admettent en privé. Lors de nos discussions privées à la NSF, ils ont admis que la solution de mettre un gigantesque cluster Linux homogène tout IBM tout PC sur 4 sites (la dernière terascale facility) n'était pas forcément optimale du point de vue technique.

Si l'Europe avait une très forte industrie informatique, les choses se passeraient probablement de la même façon. Aujourd'hui, nos choix stratégiques à l'échelle des gros projets nationaux sont nettement plus indépendants qu'aux USA.

ANNEXE A Abréviations

ACI GRID : Action concertée incitative Globalisation des ressources informatiques et des données

Akenti : architecture et modèle de sécurité (<http://www-itg.lbl.gov/Akenti/>)

CCA : Common Component Architecture (<http://www.cca-forum.org>)

CML : Chemical Markup Language (<http://www.xml-cml.org>)

Condor : Gestion distribuée de jobs

Condor-G : version de Condor utilisant Globus

GAA : Generic Authentication and Authorization

(<http://www.ietf.org/proceedings/98dec/slides/cat-linn-98dec/>)

GGF : Global Grid Forum

GSI : Grid Security Interface (<http://www.globus.org/security/>)

GSI-FTP : FTP fondé sur GSI

IETF : Internet Engineering Task force (<http://www.ietf.org>)

IRTF : Internet Research Task Force (<http://www.irtf.org>)

RNRT : Réseau national de recherche en télécommunication

RNTL : Réseau national de recherche en technologie logicielle

SOAP : Simple Object Access Protocol (<http://www.w3.org/TR/SOAP>)

TLS : IETF Transport Layer Security (<http://www.ietf.org/html.charters/tls-charter.html>)

UDDI : Universal Description, Discovery and Integration (<http://www.uddi.org>)

WebDAV : Web-based Distributed Authoring and Versioning (www.webdav.org)

XML : Extensible Markup Language

WSDL : Web Services Description Language (<http://www.w3.org/TR/wsdl>)

ANNEXE B Etapes du réseau vBNS (et vBNS+)

1995

- April 1995 - Start of 5-year cooperative agreement between the NSF and MCI to build and run vBNS
- World's first IP/ATM OC-3 (155 Mbps) nationwide backbone
- vBNS connects five Super Computer Centers (SCCs)
- 1st government customer - NSF
- End-to-end performance testing demonstrates 120 Mbps
- Traffic collection at SCCs using NNStat
- IP Multicast with Mbone tunneling across SCCs
- MAE East, PacBell, Ameritech, Sprint NAP connectivity established
- Cisco 7000-series routers used as main IP routers
- Netstar Gigarouters installed with HIPPI interfaces to connect SCCs

1996

- Native IP Multicast using PIM-DM (Protocol Independent Multicast - Dense Mode)
- Prototype OCxMON deployed - OC-3 level passive IP traffic monitoring
- Early Packet discard evaluations performed on Fore ATM switches
- OC-12 test network established between San Diego and Pittsburgh SCCs
- vBNS develops class-based queuing software to work on Netstar Gigarouters ATM interfaces
- UNI 3.1 ATM signaling with SVC capability deployed
- Peering established with NASA's Internet Sciences Network (NISN)
- Fore ASX1000s and Cisco 7507 platforms replace original equipment in preparation for backbone upgrade
- NSF announces High-Performance Connections Program (HPC) soliciting to fund High-Speed connections (DS-3 and above) for university connections to the vBNS
- Backbone latency measurement system established to record Round Trip Times (RTT) across all backbone paths

1997

- IPv6 Tunnels established across vBNS
- vBNS publishes the 1st study of Internet Traffic on an IP Backbone, Wide-Area Internet Traffic Patterns and Characteristics (IEEE Network Nov/Dec issue)
- Performance hosts deployed, OC-12 ATM attached, deployed at all backbone terminals running full test suite
- Peering established with DREN
- Backbone upgraded to OC-12
- End-to-end performance testing across backbone yields TCP throughput exceeding 500 Mbps
- Peering established with IDREN (Interim DREN), ESNET, DARTNET
- vBNS backbone expands to include nine MCI terminals (POPs) across the US
- OC3MONs deployed at all sites
- First access-connections (NSCA and SDSC) upgraded to OC-12
- ATM VBR mesh added to existing UBR to support future QoS services
- Peering established with international R&E (Research and Education) networks
- Approximately 30 universities connected at DS-3 and OC-3 as part of the HPC program

1998

- Native IPv6 service offered with dedicated IPv6 routers (Cisco 4700s)
- IP Multicast backbone configuration upgraded to PIM-SM
- Multicast BGP introduced
- Internet 2 forms with vBNS as initial backbone
- Upgrade to OC-12 POS begins
- Intensive equipment evaluations performed on both Cisco GSRs and Juniper M40 routers in competition to select a next generation backbone network routers platform
- Junipers M40s selected as backbone routers
- First fielded Juniper router in a production network
- vBNS publishes, Performance Measurement on the vBNS (Proceedings of INTEROP Engineering Conference May)
- vBNS publishes Technology Overview paper, vBNS: Not Your Father's Internet (IEEE Spectrum July issue)
- vBNS publishes, Packet-based Approaches to ATM Cell Policing, and their Effects on Internet Traffic (IEEE Communications)
- Commitment made by vBNS engineering to support IP multicast technology. Drove deployment with hands-on engineering support leading to IP multicast infrastructures across more
 - than 40 networks and institutions
- Reserved bandwidth and Weighted Fair Queuing (WFQ) tested on new Cisco platforms in vBNS lab as part of continuing QoS development
- OC3MONs upgraded to OC12MONs for OC-12 line-rate backbone flow monitoring
- ATM PNNI trials on vBNS test bed
- Separate QoS test bed established between vBNS engineering lab in Reston, VA, and Pittsburgh SCC
- MPLS trials begin (in cooperation with Georgia Tech)

1999

- 100+ Universities and R&E networks connected to vBNS
- NIH added to vBNS
- First OC-48 Packet Over SONET (POS) installed on vBNS backbone
- ATM PNNI deployed on vBNS
- MPLS interoperability testing between Cisco and Juniper platforms
- MSDP, an interdomain multicast protocol, deployed on vBNS
- Multicast interoperability testing between Cisco and Juniper platforms
- Introduction of Multicast Source Discovery Protocol (MSDP) on vBNS
- vBNS based on dual backbone topology with nationwide OC-12 POS and ATM trunks
- IPv6 production address space (a sub-TLA) Delegated from ARIN
- Juniper M40 routers installed across vBNS
- Largest heterogeneous deployment (using multiple routers - Cisco and Juniper) of native IP multicast
- Testing of terabit-class router (Avici TSR) begins
- Juniper M40 becomes primary vBNS backbone router
- High-bandwidth flow reports (derived from OCxMON data) published on a regular basis on vBNS Web page
- Over 20 MSDP Peers established
- MCI WorldCom demonstrates its own Policy Server at the Internet2 Qbone Bandwidth Broker Operability event
- Deployed NTP (Network Time Protocol) - based system for one-way delay measurements across vBNS

2000

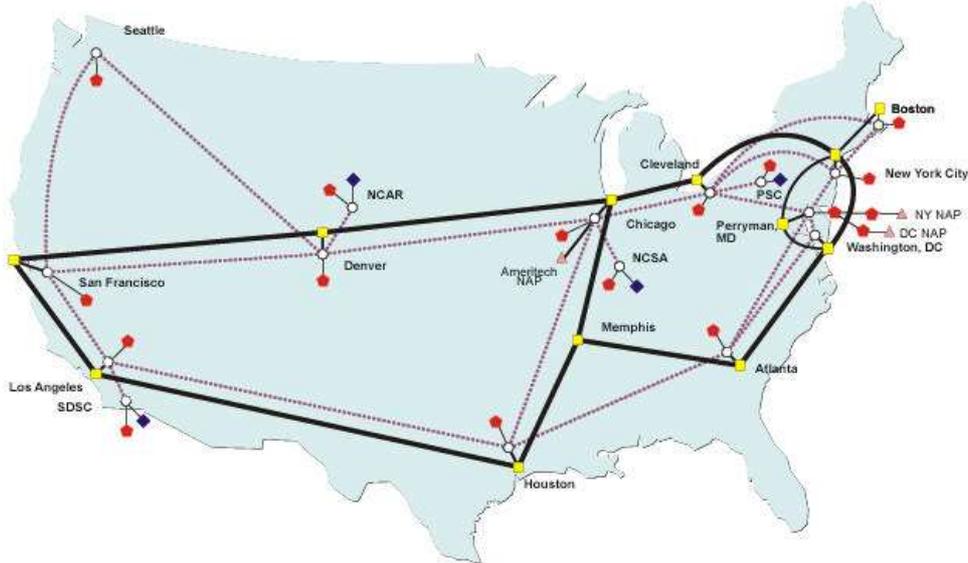
- vBNS begins upgrade to OC-48 POS Backbone
- Multi-protocol Label Switching (MPLS) deployed across vBNS - First known MPLS deployment in the US
- Constrained Shortest Path First (CSPF) configured in combination with primary and secondary physically diverse label switched paths (LSPs), resulting in a technically superior backbone in speed, performance and restoration over any ATM-centric technology
- Commercial offering of vBNS+ begins
- Trials of VoIP technology conducted
- Patented high performance (>100 Mbps) multicast test deployed and activated
- Content Distribution platforms (Cacheflows) added to network
- Prototype OC12MON developed for POS
- SLA established promising 0.001% Packet loss and 100% availability
- Active performance system refined to IPERF based measurements
- vBNS+ releases an RFI (request for information) document detailing WorldCom's requirements for a next-generation QoS and multicast -capable edge router
- Advanced test suite developed to include documented regression testing applied to all new router code releases for both unicast and multicast testing

2001

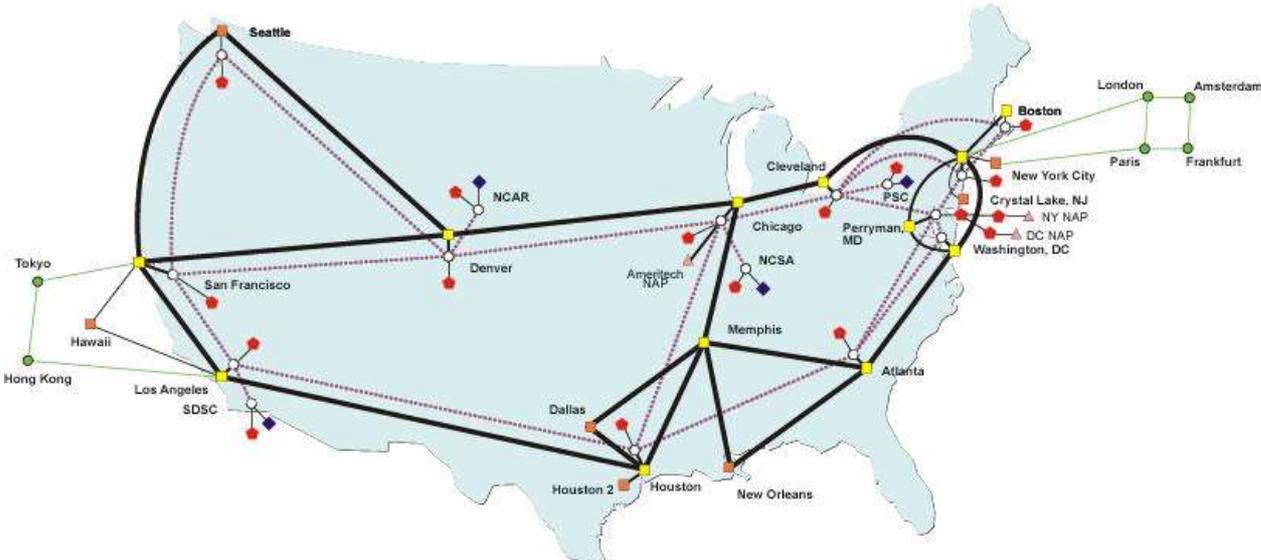
- Line rate filtering at flow-level-granularity capability established on vBNS+ Juniper M40 routers
- High speed packet encrypters evaluated
- Prototype Customer Traffic Profiling capability
- Customer Traffic Profiling (CTP) / Customer-centric Network Information Console (CNIC) services introduced
- Combination of passive measurement capabilities (SNMP/OCxMON/Commercial Flow Monitor) and active measurement systems (IPERF, Patented Multicast Test, RTT, NTP) established as the world's most instrumented network*
- Trials begin for a next-generation QoS edge router
- vBNS publishes, Multicast Performance Measurement on a High-Performance IP Backbone (Elsevier Computer Communications Journal, early 2001)
- CNIC introduction provides customers with an unprecedented level of network usage insight
- Commercial flow monitors deployed on vBNS+
- Prototype OCxMON-based core backbone Intrusion Detection System (IDS) developed

ANNEXE C Cartes du réseau vBNS/vBNS+

Etat actuel

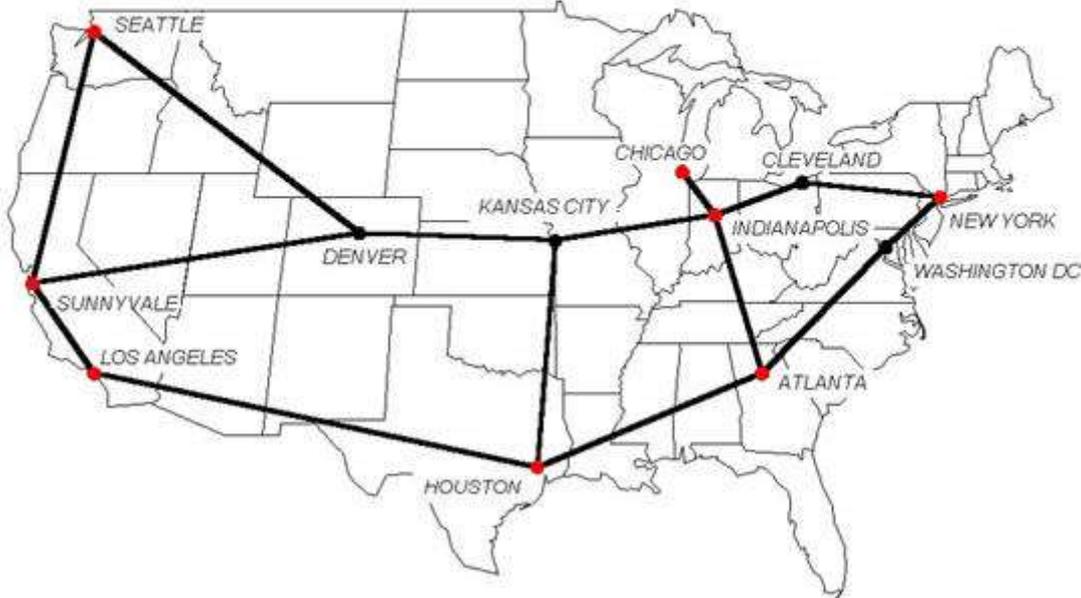


Perspectives d'évolution



ANNEXE D Cartes du réseau Abilene

ABILENE NETWORK - OCTOBER, 2001



ANNEXE F Projets financés par le DoE

Financement via le programme SciDAC :

- trois projets coopératifs (impliquant majoritairement les grands laboratoires du DoE) (\$6.81M)
 - « A National Collaboratory to Advance the Science of High Temperature Plasma Physics for Magnetic Fusion »
Participants : General Atomics, Princeton Univ. Univ of Utah, PPPL, ANL, LBNL.
Technologies : GLOBUS
 - « Particle Physics Data Grid Collaborative Pilot »
Participants : Univ of Wisconsin, UCSD, Caltech, BNL, FNAL, Jlab, ANL, LBNL, SLAC
Technologies : GLOBUS, Condor
 - Earth Systems Grid II : Turning Climate Datasets into Community Resources »
Participants : ANL, UCAR, ORNL, LLNL
Technologies : GLOBUS
- un projet pilote pour le développement d'une infrastructure de services (\$1.88M) :
 - « DoE Science Grid : Enabling and Deploying the SciDAC Collaboratory Software Environment »
Participants : ANL, LBNL, PNNL, ORNL
Technologies : GLOBUS
- deux projets de recherche en middleware (\$930k)
 - « Middleware Technology to Support Science Portals »
Participant : Indiana Univ. (Dennis Gannon)
Technologies : CCA
 - « A High Performance Data Grid Toolkit : Enabling Technologies for Wide Area Data Intensive Applications »
Participants : ANL (Ian Foster), Univ. of Wisconsin, USC
Technologies : GLOBUS
- quatre projets de recherche en réseaux (\$2.5M)
 - « Security and Policy for Group Collaboration »
Participants : ANL (S. Tuecke), Univ. of Wisconsin, USC
 - « Optimizing Performance and Enhancing Functionality of Distributed Applications Using Logistical Networking »
Participant : Univ. of Tennessee (M. Beck)
 - « Bandwidth Estimation : Measurement Methodologies and Applications »
Participants : UCSD (K. Claffy), Wellesley College, Univ of Delaware
 - « INCITE : Edge-Based Traffic Processing and Service for High-Performance Networks »
Participants : Rice Univ (R. Baraniuk). LANL, SLAC
- sept centres pour l'intégration d'infrastructures logicielles (\$19.5M)
 - « Scalable Tools for Large Clusters; Resource Management; System Interfaces; System Management Tool Framework »
Participants : ORNL (Al Geist), ANL, LBNL, LANL, PNNL, SNL, Univ of Illinois
 - « High-End Computer Systems Performance: Science and Engineering »
Participants : LBNL (David Bailey), ANL, LLNL, ORNL, Univ of Illinois, Univ of Maryland, Univ of Tennessee, Univ of California
 - « Center for Component Technology for Terascale Simulation Software »

Participants : SNL (Rob Armstrong), ANL, LANL, LLNL, ORNL, PNNL, Indiana University, University of Utah

Technologies : CCA

- « Scientific Data Management Enabling Technology Center »
Participants : ANL (Ari Shoshani), LBNL, LLNL, ORNL, Georgia Inst. Of Tech., North Carolina State University, Northwestern University, University of California San Diego
- « Terascale Optimal PDE Simulations (TOPS) »
Participants : Old Dominion Univ (David Keyes), ANL, LBNL, LLNL, University of California at Berkeley, CMU, University of Colorado,
- Terascale Simulation Tools & Technologies Center
Participants : BNL (Jimm Glimm), ANL, LLNL, ORNL, PNNL, SNL, Rensselaer Polytechnic Institute, State University of New York.
- Algorithmic and Software Framework for Applied Partial Differential Equations
Participants : LBNL (Phil Colella), LLNL, Univ. of California, New York Univ. Univ. of North Carolina, Univ. of Washington, Univ. of Wisconsin

A coté de l'initiative SciDAC, le DoE finance également plusieurs projets dans le cadre d'un financement de base (\$5.5M)

- un projet coopératif
 - « Collaboratory for Multi-Scale Chemical Science »
Participants : SNL (J. Hewson), PNNL, ANL, LANL, LLNL, MIT, NIST, UCB
Technologies : outils de collaboration, groupes de discussion, ...
- sept projets middleware :
 - « CoG Kits : Enabling Middleware for Designing Science Applications, Web Portals and Problem Solving Environment »
Participants : ANL (G. von Laszewski), LBNL
Technologies : GLOBUS
 - « Storage Resource Management for Data Grid Applications »
Participants : LBNL (A. Shoshani), FNAL
Technologies : développement d'une technologie de gestion de données à grande échelle
 - « Middleware to Support Group to Group Collaboration »
Participants : ANL (R. Stevens)
 - « Pervasive Collaborative Computing Environment »
Participants : LBNL (D. Agarwal), Univ. of Wisconsin
Technologies : WebDAV, GSI FTP, SOAP, Condor-G, GLOBUS
 - « Scientific Annotation Middleware »
Participants : PNNL (J. Myers), ORNL
Technologies : XML, CML, WebDAV
 - « Reliable and Secure Group Communication »
Participants : LBNL (D. Agarwal)
 - « Distributed Security Architectures »
Participants : LBNL (M. Thompson)
Technologies : TLS, GSI, GAA, WebDAV, Akenti
- six financements dans le domaine des réseaux
 - « Self-Configuring Network Monitoring »
Participants : LBNL (D. Agarwal)
 - « The Net100 Project : Development of Network-Aware Operating Systems »
Participants : CMU (G. Huntton), ORNL, LBNL
 - « High-Performance Transport Protocols »
Participants : LANL (Wu-Feng)

- « Stability Modelling and Control of Transport Protocols (TCP) for High-Speed Data Grids »
Participants : ORNL (R. Baraniuk)
- « IQEcho : Interoperability and Quality of Service Across Heterogeneous Hardware and Software Platforms »
Participants : Georgia Tech (K. Schwan)
- « Pushing the Network Simulation Envelop – Extending the SSF to Tera-scale Problems »
Participants : ORNL (B. Wing)