

# Le réseau dans un projet de grille de calcul : l'exemple d'EGEE

Mathieu Goutelle

CNRS UREC

mathieu.goutelle@urec.cnrs.fr

Jean-Paul Gautier

CNRS UREC

jean-paul.gautier@urec.cnrs.fr

## Résumé

Dans cet article, nous vous proposons de découvrir les activités réseau d'un projet de grille : EGEE. Ces activités sont centrées autour de l'accès aux services fournis (réservation de ressources) et de la contrainte opérationnelle d'une telle architecture (interface opérationnelle avec les réseaux, institution de SLA, surveillance). Même si ce projet présente certaines spécificités, de nombreuses problématiques sont généralisables, notamment celles concernant le réseau. Elles préfigurent de l'évolution possible de l'utilisation des réseaux dans le futur, non pas seulement par EGEE ni même par d'autres projets de calcul distribué, mais par tous les utilisateurs.

## Mots clefs

grille, interface opérationnelle, *Service Level Agreement* (SLA), réservation de ressources, *monitoring*, bout-en-bout

## Glossaire

**BAR** *Bandwidth Allocation and Reservation* — Logiciel développé par l'activité JRA4 servant d'interface avec les services réseau.

**CERN** Organisation Européenne pour la Recherche Nucléaire — <http://www.cern.ch>

**ENOC** EGEE *Network Operation Centre* — Équipe en charge, dans EGEE, des problèmes opérationnels liés au réseau.

**GÉANT** le réseau GÉANT — Réseau d'interconnexion des NREN en Europe. Voir GÉANT2.

**GÉANT2** Projet européen pour construire la nouvelle génération du réseau d'interconnexion entre les *National Research and Educational Network* (NREN) (<http://www.geant2.net/>).

**GGF** *Global Grid Forum* — Le GGF représente, pour le domaine des grilles de calcul, un organe d'édition de normes et d'élaboration d'éléments permettant le travail en commun (<http://www.gridforum.org>).

**g-Lite** *Middleware* de grille développé dans le cadre du projet EGEE. Il intègre de nombreux composants provenant d'autres *middlewares* (Alien, LCG, etc.).

**JRA** *Joint Research Activity*

**L-NSAP** *Local Network Service Access Point* — Interface fournie par JRA4 pour la configuration des routeurs d'extrémité (dans les sites).

**LHC** *Large Hadron Collider* — Le grand collisionneur de hadrons est un accélérateur de particules à très hautes énergies (<http://lhcb.web.cern.ch>).

**MPLS** *Multi-Protocol Label Switching* — Introduit des mécanismes de type circuit dans un réseau de paquets type IP.

**NM-WG** *Network Measurement Working Group* — Groupe de travail du GGF consacré à l'identification de métriques réseau utiles pour la grille et à l'élaboration de mécanismes standardisés pour la publication et l'accès à ces données (<http://nmwg.internet2.edu>).

**NOC** *Network Operation Centre* — Équipe en charge de la gestion et de la maintenance d'un réseau.

**NPM** *Network Performance Monitoring* — Outil d'accès à des données de surveillance du réseau.

**NREN** *National Research and Educational Network* — Réseau National pour la Recherche et l'Éducation

**NSAP** *Network Service Access Point* — Interface fournie par un NREN pour réserver et utiliser un service réseau ; dialogue avec BAR.

**PERT** *Performance Enhancement and Response Team* — Groupe d'experts constitués dans le cadre de GÉANT2 pour répondre aux problèmes de performances des utilisateurs.

**SA** *Service Activity*

**SLA** *Service Level Agreement* — « Contrat » passé entre un client et un fournisseur de services et spécifiant le niveau de qualité du dit-service.

**SLR** *Service Level Requirement* — Description des besoins d'un utilisateur.

**SLS** *Service Level Specification* — Description des caractéristiques techniques d'un service en fonction des besoins exprimés dans un SLR.

**TNLC** *Technical Network Liaison Committee* — Comité interne à EGEE servant d'interface entre EGEE et les NREN pour toute question technique.

# 1 Introduction

Le projet EGEE (*Enabling Grid for E-sciencE*<sup>1</sup>) est un projet financé par l'Union Européenne. Il rassemble les experts de 70 partenaires, académiques et industriels, répartis dans 27 pays afin de bénéficier au mieux de l'expérience accumulée dans les projets nationaux et internationaux de grille de calcul. Le but du projet EGEE est de faire évoluer l'infrastructure de grille de calcul préexistante vers un fonctionnement opérationnel afin de fournir aux utilisateurs une disponibilité des ressources 24h/24 en Europe. Nous nous proposons de montrer à travers l'exemple d'EGEE l'impact des grilles de calcul opérationnelles sur les activités réseaux dans une perspective de services de bout en bout.

Après une courte présentation du projet (chapitre 2), nous allons détailler au chapitre 3 quelles sont les activités relatives au réseau dans EGEE. Ensuite, au chapitre 4, nous décrirons les implications que ce type de projet peuvent avoir sur les réseaux, notamment les réseaux (locaux, métropolitains, régionaux, *etc.*) situés entre le NREN et le centre de ressources (calcul, stockage). Enfin, nous concluons au chapitre 5.

## 2 Le projet EGEE

Le projet EGEE a commencé en avril 2004 pour une période de deux fois deux ans et fait suite à plusieurs projets de grilles de calcul, notamment DataGrid<sup>2</sup>. Le but est de proposer à un utilisateur des ressources de calcul et de stockage très importantes à travers une interface simplifiée et des services d'authentification, de soumission des tâches et de gestion des données (dépôt, réplique, récupération, *etc.*). Les ressources sont fournies au niveau national, par les différents partenaires du projet, charge à celui-ci d'« unifier » ces ressources en une infrastructure homogène. À cette date, EGEE regroupe plus de 100 sites en Europe et dans le monde dans une vingtaine de pays, fournissant plus de 10 000 processeurs et plus de 5 péta-octets<sup>3</sup> de stockage. Le *middleware* utilisé dans EGEE, g-Lite, intègre des composants de *middlewares* existants (LCG, Alien, EDG, *etc.*) et propose de nombreuses améliorations ou nouveautés.

Originellement prévue pour les expériences de physique nucléaire du CERN et le traitement des énormes quantités de données générées par le *Large Hadron Collider* (LHC), le nouvel accélérateur de particules du Organisation Européenne pour la Recherche Nucléaire (CERN), la grille accueille désormais d'autres applications. Citons par exemple les applications biomédicales (simulations de traitement tomographique, aide à la décision, analyse de séquences ADN de protéines), des Sciences de la Terre (traitement de données sismiques, observation de la Terre, climatologie, *etc.*) ou de chimie (étude de la chimie par modélisation numérique par exemple).

<sup>1</sup><http://www.eu-egee.org>

<sup>2</sup><http://eu-datagrid.web.cern.ch/eu-datagrid/>

<sup>3</sup>Un péta-octet correspond à 1000 téraoctets ( $1 \times 10^{12}$  octets).

Pour interconnecter les nombreux centres de ressources, cette grille dépend fortement du réseau sous-jacent, de ses performances et des services qu'il peut fournir. Elle s'appuie sur les réseaux nationaux de la recherche et de l'éducation, les NREN, et sur GÉANT, le réseau européen d'interconnexion des NREN pour construire le « réseau EGEE », réseau virtuel interconnectant les nœuds EGEE entre-eux. C'est ce réseau, impliquant une quarantaine de NREN en Europe et dans le monde, avec ces hétérogénéités de structure et de connectivité dont il est question dans le cadre d'EGEE. Le réseau EGEE profitera des services fournis et développés dans le cadre de ces réseaux. Pour ce faire, deux activités « réseau » ont été prévues dans EGEE. La première, *Network Resource Provision* (SA2) concerne la partie opérationnelle. L'activité de développement de services réseau à l'intérieur d'EGEE est assurée par l'activité JRA4, *Development of Network Services*. L'UREC est impliquée entre autres dans chacune de ces deux activités.

## 3 Les activités réseau dans EGEE

D'un point de vue général, notre but est de faire reconnaître le réseau comme une **ressource** de la grille, tout comme les ressources de calcul et de stockage. À l'heure actuelle, le réseau reste assez « transparent » et n'est pas considéré à part entière comme faisant partie de ce qu'il faut gérer dans une telle infrastructure. Deux raisons principales peuvent expliquer cela :

- La première est que le réseau est une **ressource externe** à la grille puisqu'elle est fournie par ailleurs (par les NREN aux centres de ressources par exemple), hors du contexte d'EGEE et plus généralement de la grille. Elle dépend de plus des NREN, *i.e.* d'entités administrativement hors du projet et donc hors de l'infrastructure ;
- La seconde est qu'il manque des moyens de **gérer automatiquement cette ressource** au niveau applicatif, et particulièrement les services que le réseau pourrait apporter (autre que la simple connectivité). Cela empêche de construire une architecture de gestion des ressources de la grille, en y incluant évidemment le réseau.

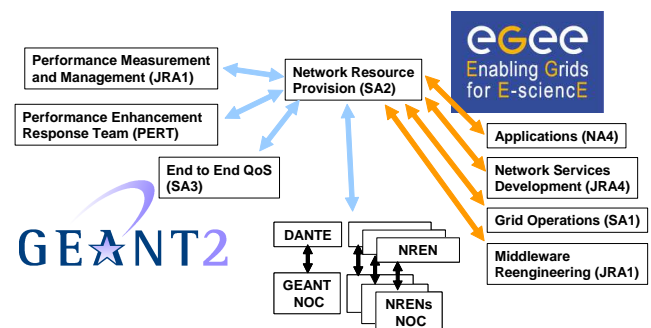


Figure 1 – Interactions entre EGEE et les réseaux

Pourtant, la grille est très dépendante des réseaux et de leurs services, sans avoir aucun contrôle réel sur cette ressource. C'est pourquoi nous travaillons en étroite collaboration avec GÉANT et les NREN (cf. figure 1) pour anticiper sur les futurs services réseau, par exemple ceux développés dans le cadre de GÉANT2, et être ainsi prêt à les utiliser dès leur disponibilité. Cela inclut par exemple la définition commune d'interfaces logicielles et de modèles opérationnels. En tant qu'interface avec les fournisseurs de réseau, nous devons travailler avec les activités d'EGEE qui sont utilisateurs des réseaux, telles que les applications ou le développement du *middleware*.

### 3.1 L'activité JRA4 : « Network Service Development »

L'activité JRA4 est une activité de développement de logiciels prévus pour être intégrés au *middleware*. Cette activité est scindée en deux sous-activités, *Bandwidth Allocation and Reservation* (BAR) et *Network Performance Monitoring* (NPM).

NPM a été développé pour accéder de façon unifiée et homogène à des données de surveillance et de performance du réseau EGEE, fournies par différents domaines et différentes structure de mesure [1]. L'approche de JRA4 a donc été de compter sur une standardisation des interfaces d'accès aux données (cf. figure 2). Le choix s'est porté sur les travaux du *Global Grid Forum* (GGF) et plus particulièrement du groupe de travail *Network Measurement Working Group* (NM-WG).

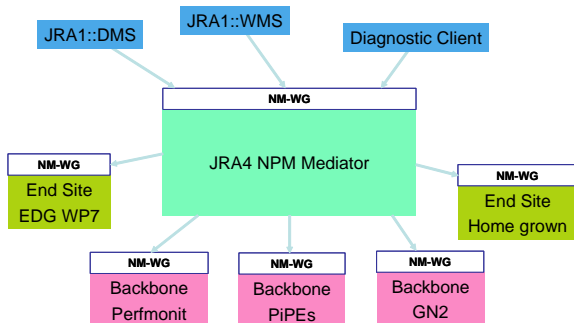


Figure 2 – Principe de l'architecture de NPM

L'utilisation d'une interface standardisée permet d'une part d'accéder à tous les outils de mesure existants, à condition qu'ils fournissent la même interface pour publier les données. C'est par exemple le cas des outils de mesure de bout-en-bout développés dans le cadre de DataGrid<sup>4</sup>, mais aussi de l'infrastructure « Perfmonit » de mesure dans le cœur du réseau GÉANT<sup>5</sup>. D'autre part, cela permet à de nombreux outils (outil de diagnostic, composants du *middleware*, utilisateur) de pouvoir accéder à ces données de façon unifiée (*i.e.* sans choisir quel est l'outil approprié pour sa demande)

<sup>4</sup><http://ccwp7.in2p3.fr/>

<sup>5</sup><http://www.dante.net/server/show/nav.00100q003002>

<sup>6</sup>notamment le changement du niveau d'abstraction entre classes de services applicatives et services orientés technologie réseau.

<sup>7</sup>Presque 50% des NREN ont déployé *Premium IP* d'après une étude faite au début de l'année 2005 (<https://edms.cern.ch/document/477861>).

puisque NPM agrège les données qu'il collecte et les publie toujours à travers la même interface.

BAR, la seconde activité de JRA4, doit permettre aux applications s'exécutant sur la grille d'utiliser les services réseau en effectuant des réservations [2]. BAR doit être vu comme une interface avec les mécanismes présents dans le réseau : en d'autres termes, ce n'est pas BAR qui configure le réseau, mais un service, le *Network Service Access Point* (NSAP), sous le contrôle du NREN. Il est donc crucial pour BAR de travailler en étroite collaboration avec les réseaux et d'obtenir une participation importante des NREN, notamment à travers les activités JRA3 et SA3 du projet GÉANT2. Pour résoudre le problème de la configuration des équipements d'extrémités et des réseaux locaux des sites, il est prévu l'utilisation d'un *Local Network Service Access Point* (L-NSAP) installé dans les centres de ressource.

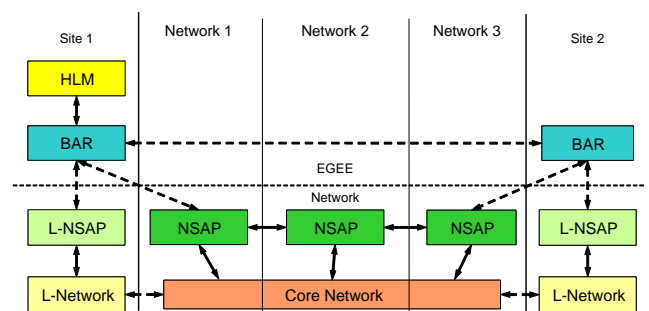


Figure 3 – Architecture de l'interface entre BAR et le réseau. HLM signifie Higher Layer Middleware.

L'architecture complète des mécanismes impliqués dans BAR est décrite à la figure 3 : une requête parvenant à BAR est d'abord validée (correction des paramètres). Elle est ensuite envoyée après traitement<sup>6</sup> aux autres composants, en premier lieu le NSAP du réseau adjacent qui va vérifier la faisabilité de la réservation demandée au niveau du réseau en dialoguant avec les NSAP impliqués dans le chemin, puis au service BAR du site distant. Les deux instances de BAR vont questionner leur L-NSAP respectif pour vérifier la faisabilité de la requête au niveau des réseaux locaux des deux sites. Si la requête est rejetée par l'un des NSAP ou L-NSAP, la requête est rejetée par BAR qui va demander l'annulation de la réservation auprès des composants ayant fournis une réponse positive. Sinon, la configuration des réseaux est réalisée par chacun des NSAP et L-NSAP impliqués.

Pour le moment, le seul service réseau qui pourra être utilisé par BAR est le service *Premium IP* parce qu'il est le plus déployé dans les NREN<sup>7</sup>. Les applications d'EGEE auront la possibilité d'utiliser d'autres services réseau, dès leur disponibilité via des requêtes à un NSAP.

JRA4 est aussi en charge d'évaluer l'impact du déploiement de IPv6 dans une infrastructure de grille telle que celle d'EGEE [3]. Ce travail a identifié les avantages que IPv6

pourrait apporter à l'infrastructure et aux applications, mais aussi les problématiques et les spécificités de migration et de coexistence entre IPv4 et IPv6 et l'intégration de IPv6 dans le *middleware*.

### 3.2 L'activité SA2 : « Network Resource Provision »

Si JRA4 peut être vu comme l'activité en charge de « l'interface logicielle » entre la grille et le réseau, SA2 en assure « l'interface opérationnelle ». Cette relation avec les NREN et GÉANT a été par exemple formalisée par la création du *Technical Network Liaison Committee* (TNLC). Ce comité permet de présenter les travaux de chacun à l'autre communauté et de partager des informations.

Cette interface opérationnelle est aussi le lieu des échanges d'informations à propos des incidents, des maintenances et des problèmes rencontrés par les utilisateurs sur le réseau EGEE. Pour faciliter ce dialogue, des procédures ont été définies [4] pour les échanges de « tickets » entre les systèmes de gestion d'incidents des *Network Operation Centre* (NOC) des NREN et celui d'EGEE<sup>8</sup>. Ces échanges, indispensables pour bâtir une grille opérationnelle en tenant compte du réseau, sont bidirectionnels : un utilisateur de la grille peut signaler au support d'EGEE un problème lié au réseau. Celui-ci va être assigné à l'EGEE *Network Operation Centre* (ENOC), qui va alors faire un diagnostic pour localiser le problème (site, machine, *etc.*), si besoin tenter d'identifier le réseau concerné et le contacter le cas échéant. Il suivra ensuite l'évolution du ticket, prenant en charge les demandes de la part du NOC ou de l'utilisateur. Compte tenu du nombre important à la fois d'utilisateurs et de réseaux en jeu, il serait difficile de les faire dialoguer directement entre eux. C'est pourquoi la création de l'ENOC nous a semblé nécessaire, comme point de contact unique pour les parties impliquées (centres de ressources, NOC et utilisateur).

En tant qu'interface opérationnelle entre le réseau et la grille, SA2 doit aussi veiller à ce que les besoins des applications en terme de réseau soient satisfaits autant que possible. Cela passe d'abord par la collecte des besoins spécifiques, dérivés des cas d'utilisations des applications. À partir de ces besoins, nous avons identifié un certain nombre de classes de services [5] agrégeant des utilisations semblables (trafic de messages ou de flux interactifs, transferts « lourds », réseaux privés, *etc.*). À ces classes de services, indépendantes des technologies sous-jacentes, doivent être associées des services fournis par les réseaux [6] : par exemple, un service IP comme *Premium IP*<sup>9</sup> peut être une possibilité pour un trafic de messages interactifs. Cette traduction des classes de service en services réseau doit être effectuée dans le *middleware* afin que les requêtes à un NSAP soit faite en termes de services réseau. Cette correspondance pourra évoluer en fonction des besoins des applications et de l'apparition de nouveaux services réseau.

Si l'utilisation de services réseau se généralise, il est indispensable pour les fournisseurs de gérer soigneusement la ressource qu'ils mettent à disposition de leurs clients. Pour cela, le passage par la « signature » d'un SLA semble nécessaire. En effet, afin d'être à même de fournir aux utilisateurs d'EGEE un service, il faut être capable de décrire ces caractéristiques techniques et administratives. Le SLA est l'aboutissement d'un travail entre le fournisseur de service et son utilisateur, qui passe par l'expression d'un besoin sous forme d'un *Service Level Requirement* (SLR), l'identification et la description technique d'un service adapté à ce besoin sous forme d'un *Service Level Specification* (SLS). Le SLA n'est finalement que la formalisation de l'accord entre les deux parties pour l'utilisation d'un service décrit dans un SLR.

Dans ce domaine, la tâche de SA2 a été de proposer des modèles de SLA [7] suffisamment généraux pour tenir compte de l'hétérogénéité des situations qui peuvent exister parmi la variété de réseaux impliqués dans EGEE<sup>10</sup> mais aussi suffisamment précis pour être utilisables. Ces modèles sont constitués de deux parties. La première contient les informations administratives : contact technique, durée de la réservation, procédure en cas d'incident, *etc.* La seconde décrit les caractéristiques techniques de la réservation : source et destination concernées, garanties, limites de trafic, politique en cas d'excès, *etc.* Ils tiennent compte des diverses situations que le réseau EGEE contient en terme de déploiement et d'accessibilité de service. Plusieurs scénarios ont ainsi été élaborés afin de pallier l'hétérogénéité des réseaux (implémentation partielle ou absence d'un service dans un domaine emprunté par le chemin). Chaque SLA est défini pour un domaine particulier (un NREN), les SLA étant ensuite combinés afin de fournir un SLA de bout-en-bout qui décrira le service fourni à l'application (cf. figure 4).

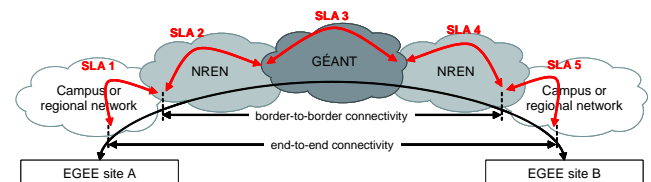


Figure 4 – Combinaison de SLA par domaine pour construire le SLA de bout-en-bout

Ce passage par les SLA est une nouveauté, à la fois pour les NREN et pour la grille. Pour cette raison et afin de gagner de l'expérience dans ce domaine et ainsi valider les procédures définies pour l'établissement des SLA, SA2 a conduit une expérimentation (cf. figure 5) permettant d'utiliser le service *Premium IP* entre plusieurs sites en France, en Grèce et en Russie. Ces sites ont été choisis pour évaluer la grande variété de situations possibles : site directement connecté à son NREN, présence de réseaux intermédiaires (métropolitain ou de campus). Cette expérimentation a conduit à la définition de SLA décrivant le service fourni dans chaque domaine impliqué dans l'expérimentation.

<sup>8</sup><http://www.ggus.org/>

<sup>9</sup>Service fournissant aux flux marqués une priorité stricte sur les autres flux, limitant ainsi le délai au maximum.

<sup>10</sup>Nous avons pu recenser plus d'une quarantaine de réseaux participant à l'interconnexion des sites d'EGEE.

La question concernant le *monitoring* des SLA reste en suspens. La solution à apporter à ce problème dépendra de la politique adoptée en matière de respect des SLA et la granularité du *monitoring* disponible : soit une surveillance des caractéristiques **en continu** pendant toute la durée de la réservation, soit une surveillance *a posteriori* pour vérifier si l'application a reçu les garanties demandées **en moyenne** sur la durée de l'utilisation.

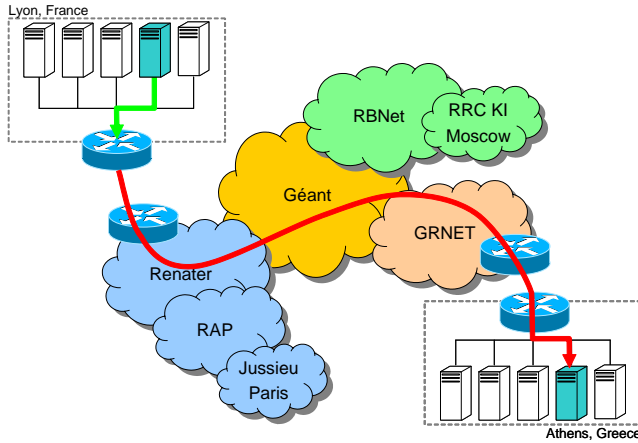


Figure 5 – Expérience de mise en place des SLA

### 3.3 Cas d'utilisation d'un service réseau

Détaillons ce que pourrait être le mécanisme utilisé par une application pour utiliser un service réseau entre deux sites. Nous allons nous placer dans la situation qui prévaut actuellement, *i.e.* celle qui impose une configuration manuelle des équipements du cœur du réseau et donc un temps d'établissement d'une réservation d'au moins un jour ouvré. Dans un avenir proche<sup>11</sup>, même si cette configuration devient automatique, l'utilisateur devra quand même prendre en compte que la configuration sera appliquée aux équipements périodiquement seulement (deux fois par jour par exemple).

Cette utilisation de service nécessite l'établissement préalable de SLA pour ce service entre les différents acteurs concernés : entre les deux sites et leur réseau adjacent et entre chaque réseau et le réseau voisin sur le chemin de données. Ces SLA peuvent être établis en amont, **pour une durée relativement longue** (un ou deux ans par exemple). Ils contiennent entre autre de nombreuses données techniques sur le niveau de qualité que l'utilisateur est en droit d'attendre de ce service.

Pour utiliser ce service, l'initiateur de la requête<sup>12</sup> va décrire ses besoins en termes qualitatifs en choisissant une des classes de services (trafic interactif, flux interactif, transfert lourd, *etc.*) et en termes quantitatifs (débit, durée, *etc.*). Lors de la soumission d'une tâche à la grille, des composants du *middleware* vont recevoir cette requête (*Service Request*) et faire les nécessaires vérifications (authentification, autorisation et gestion des ressources). Alors, si la requête est valide,

le *middleware* va dialoguer avec le réseau à travers BAR pour obtenir une réservation **à moyen terme** (quelques semaines, voire quelques mois). En combinant les SLA des domaines traversés, il est possible de construire une sorte de SLA de bout-en-bout, décrivant les caractéristiques attendues du service du point de vue de l'application. Ce sont ces caractéristiques (délai, débit, *etc.*) que le *middleware* d'EGEE devra surveiller afin de vérifier que l'application a obtenu un service conforme à sa requête ou non. Le cas échéant, l'ENOC pourra être consulté afin de diagnostiquer le problème et si besoin demander une action à l'un des NOC. Une autre solution pour l'ENOC sera de consulter le *Performance Enhancement and Response Team* (PERT) de GÉANT : cette équipe d'experts a été constituée dans le projet GÉANT2 pour justement régler les problèmes de performances de bout-en-bout.

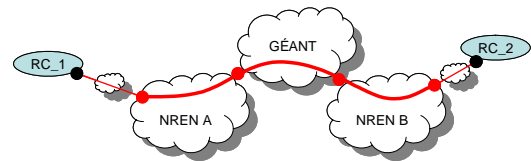


Figure 6 – Première étape de la réservation de ressource (Service Request)

À ce moment là (cf. figure 6), seul le cœur du réseau est configuré pour accepter le trafic en fonction du groupe d'émetteurs et de récepteurs possibles (préfixes d'adresses IP par exemple). Il est encore impossible de connaître les machines d'extrémités (sauf cas particulier) puisque celles-ci seront choisies ultérieurement au moment de l'allocation de ressources à la tâche. Pour éviter les utilisations « frauduleuses » ou par inadvertance du service, les équipements des sites sont encore configurés pour empêcher l'utilisation de cette réservation (nœuds en noir sur la figure 6).

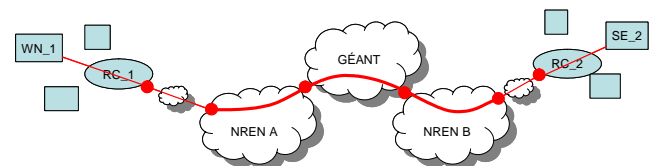


Figure 7 – Seconde étape de la réservation de ressource (Service Activation)

Lorsque la tâche arrive sur le nœud du cluster choisi par le *middleware*, il reste à activer la réservation pour la durée de cette exécution. Cette étape d'activation (ou *Service Activation*) **à court terme** va principalement permettre de configurer les équipements des sites (en rouge désormais sur la figure 7) pour autoriser les deux machines choisies pour l'exécution de la tâche à émettre du trafic dans le service réseau réservé. La ressource réservée peut être utilisée par une autre tâche soit en même temps s'il reste de la capacité disponible à l'intérieur de la réservation, soit après la fin de la

<sup>11</sup> avant la fin de GÉANT2, *i.e.* août 2008.

<sup>12</sup> un utilisateur, une application ou un composant du *middleware*.

première exécution tant que la réservation est encore valide dans le temps.

Ce mécanisme en deux étapes est indispensable pour deux raisons :

- Il permet de s’affranchir de la différence de granularité de durée de réservation entre le réseau (réservation de plusieurs jours) et la grille (réservation sur quelques minutes ou plus). Le passage à l’échelle est aussi facilité puisque de nombreuses réservations « grille » peuvent être agrégées en une réservation « réseau » et gérées ensuite par le *middleware* ;
- Il permet de résoudre le problème de l’autorisation des trafics au niveau des équipements du réseau. La configuration des équipements réseau pour l’utilisation de la réservation ne peut être finalisée qu’au moment où l’on connaît les machines d’extrémités et que l’on est sûr que celles-ci peuvent utiliser la réservation.

## 4 Les conséquences sur les réseaux

La mise en place d’une infrastructure de calcul distribué telle que la grille a des conséquences sur les réseaux. La plus évidente (et celle attendue par un administrateur) peut être une augmentation de l’utilisation des liens, même si cela est loin d’être automatique et forcément significatif. En effet, les applications « grille » n’engendrent pas toutes des transferts de gros volume de données. Le cas échéant<sup>13</sup>, des infrastructures privées sont envisagées [8] afin d’une part de garantir le service pour les expériences et de ne pas perturber l’utilisation normale des réseaux. L’Internet public ne sera alors utilisé que pour l’accès à la grille, la consultation des données ou l’exécution d’applications spécifiques.

Parmi les conséquences inattendues, citons d’abord les conséquences directes sur les réseaux de collecte jusqu’aux NREN : réseau local, métropolitain puis régional selon la hiérarchie adoptée dans chaque pays, certains pouvant être absents. La première conséquence réside dans le besoin de gestion opérationnelle. Ce n’est pas tant la haute disponibilité des réseaux **de bout-en-bout** qui est critique, l’architecture distribuée de la grille s’accommodant relativement bien de l’indisponibilité d’un site. C’est plutôt la remontée et la résolution des incidents qui sont nécessaires. La disponibilité des informations dans un système centralisé à destination des utilisateurs de la grille est un besoin qui a un impact fort sur les réseaux s’ils n’ont pas déjà un système opérationnel en place.

La seconde conséquence importante de l’arrivée d’un nœud de la grille sur un réseau est le besoin de service de bout-en-bout, que ce soit en terme de dimensionnement (bande passante disponible), de configuration (autorisation de flux spéciaux) ou en terme de qualité. Atteindre le « bout-en-

bout » est une volonté mais présente un véritable défi si l’on veut dépasser le *best-effort* IP et fournir par exemple de la qualité de service IP ou encore des tunnels basés sur *Multi-Protocol Label Switching* (MPLS). Pour évaluer la difficulté de la tâche, il suffit même de se restreindre à la disponibilité d’une bande-passante (même non garantie) de 1 Gbit/s entre deux nœuds de stockage ! Si cela est sans doute possible à réaliser entre deux centres de ressource importants, cela relève, dans la majorité des cas, du parcours du combattant (contacts multiples, synchronisation des configurations, accord sur les SLA *etc.*), sans compter les nombreuses impossibilités techniques (matériel inadapté, incompatibilité, ressource indisponible, *etc.*). Généraliser la mise à disposition de ce genre de services à une grande échelle semble de toute façon irréaliste si tout le travail préalable conserve une trop grande partie d’intervention manuelle.

C’est pourquoi une vision à long terme de l’évolution des réseaux penche vers l’apparition de mécanismes offrant à l’utilisateur un contrôle plus important de la ressource réseau. Ne pensons pas uniquement aux utilisateurs de la grille, ceux-ci n’étant que des « pionniers », mais à tous les utilisateurs. Une partie de la gestion du réseau va à terme glisser vers cet utilisateur qui sera en mesure de demander un service au réseau qui le lui fournira (ou pas) en retour, en fonction de critères convenus entre lui et son fournisseur, des caractéristiques de la requête et des conditions sur le réseau. Ceci se fera de manière dynamique, à la demande, et surtout, donnera accès à un service de bout-en-bout et multi-domaines. De nombreuses avancées sont à prévoir, soit par l’intermédiaire du projet GÉANT2 qui prépare pour les NREN ce type de service dans les activités GÉANT2-SA3<sup>14</sup> et GÉANT2-JRA3<sup>15</sup>, soit par la volonté des équipementiers, opérateurs de télécommunications et fournisseurs d’accès via des initiatives tel que le forum IPSphere<sup>16</sup>.

Ce regroupement de la majorité des acteurs du marché des réseaux informatiques incite à croire à l’avenir de ce genre de construction. Ces projets sont d’autant plus porteurs qu’ils s’appuient sur des technologies éprouvées dans les réseaux, par exemple *Premium IP* pour GÉANT2-SA3 ou MPLS<sup>17</sup> pour GÉANT2-JRA3 et IPSphere. Finalement, ils se proposent « simplement » d’une part d’ajouter à ces mécanismes existants les interfaces de réservation à destination des utilisateurs et d’autre part de traiter la problématique de bout-en-bout, *i.e.* de faire dialoguer des domaines entre eux pour finaliser la réservation quelque soit le chemin intermédiaire. C’est là où le choix de technologies et de concepts éprouvés prend tout son sens : il est d’autant plus facile de construire une solution de bout-en-bout que les mécanismes sous-jacents sont déjà majoritairement déployés dans les domaines concernés.

L’interopérabilité des réseaux pour la fourniture de service semble de toute façon une avancée pour les utilisateurs, mais

<sup>13</sup>C’est notamment le cas dans le cadre du projet LHC.

<sup>14</sup><http://www.geant2.net/server/show/nav.893>

<sup>15</sup><http://www.geant2.net/server/show/nav.756>

<sup>16</sup><http://www.ipsphereforum.org/>

<sup>17</sup>Plus exactement GMPLS, son extension aux nouveaux protocoles de transport de niveau liaison comme les chemins optiques. Voir [9] et [10].

aussi pour les réseaux puisqu'elle simplifie et accélère les procédures de mise en place d'un service et permet de fournir à ses utilisateurs (ou ses clients) des services avancés sans complication future et impossibilité de déploiement à grande échelle.

## 5 Conclusion

Le projet EGEE est un projet ambitieux : fournir aux utilisateurs de la grille la possibilité d'utiliser la plus grande infrastructure de calcul distribué avec une qualité opérationnelle. Il est cependant réaliste : la volonté de s'appuyer fortement sur les logiciels existants et de se concentrer sur les fonctionnalités indispensables à son but ont contribué aux succès du projet lors des deux premières années.

Le réseau entre dans cette logique de qualité opérationnelle, plutôt en terme de services et de flux d'informations opérationnelles qu'en terme de qualité et de haute disponibilité, propriétés assez répandues dans la majorité des réseaux actuels. C'est pourquoi les activités réseau dans EGEE se concentrent d'une part autour de l'interface opérationnelle avec les réseaux et de la surveillance des caractéristiques et d'autre part autour des services de bout-en-bout (SLA et réservations). Ces activités travaillent pour le futur, ne serait-ce que parce qu'il manque encore certaines briques indispensables, particulièrement pour la réservation de services à la demande. Néanmoins, les voies tracées en la matière par des projets contemporains d'EGEE sur l'évolution des réseaux nous portent à penser que ces briques seront bientôt disponibles et pas uniquement dans le monde des réseaux académiques.

## Références

- [1] JRA4 activity. Definition of standardized network measurement query/response interfaces. Rapport technique DJRA4.2, EGEE – Enabling Grid for E-science, Janvier 2005. <https://edms.cern.ch/document/533215>.
- [2] JRA4 activity. Specification of interfaces for bandwidth reservation service. Rapport technique DJRA4.1, EGEE – Enabling Grid for E-science, Novembre 2004. <https://edms.cern.ch/document/501154>.
- [3] JRA4 activity. Report on implications of IPv6 usage for EGEE Grid. Rapport technique DJRA4.3, EGEE – Enabling Grid for E-science, Septembre 2005. <https://edms.cern.ch/document/603955>.
- [4] SA2 activity. Operational interface between EGEE and GÉANT/NRENS. Rapport technique MSA2.3, EGEE – Enabling Grid for E-science, Mars 2005. <https://edms.cern.ch/document/565449>.
- [5] SA2 and JRA4 activities. Survey of pilot applications requirements on networks, initial SLRs and service

classes. Rapport technique DSA2.1, EGEE – Enabling Grid for E-science, Octobre 2004. <https://edms.cern.ch/document/495204>.

- [6] SA2 activity. Initial requirements aggregation model, specification of services as SLSs on the networks. Rapport technique MSA2.2, EGEE – Enabling Grid for E-science, Janvier 2005. <https://edms.cern.ch/document/509455>.
- [7] SA2 activity. Institution of SLAs and appropriate policies. Rapport technique DSA2.2, EGEE – Enabling Grid for E-science, Mars 2005. <https://edms.cern.ch/document/565447>.
- [8] David Foster, Erik-Jan Bos, Edoardo Martelli, et Paolo Moroni. LHC high-level network architecture. Rapport technique, LHC Computing Grid project, Juin 2005. <http://lcg.web.cern.ch/LCG/PEB/gdb/nw-grp.htm>.
- [9] P. Ashwood-Smith et L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions, Janvier 2003. RFC3472.
- [10] L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling - Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions, Janvier 2003. RFC3473.

