

Les classes de service sur le Réseau Académique Parisien

De la mise en œuvre technique aux accords de niveau de service

Laurent Gydé
Réseau Académique Parisien
Laurent.Gyde@rap.prd.fr

Lionel David
Réseau Académique Parisien
Lionel.David@rap.prd.fr

Résumé

Depuis mi-2005, le Réseau Académique Parisien met en oeuvre des classes de service adaptées aux différents besoins des utilisateurs et compatibles avec les services similaires sur les réseaux de recherche nationaux et européens. Au-delà de la seule configuration des mécanismes de qualité de service sur les équipements actifs, cet article présente globalement la démarche de construction du service, de l'expression du besoin à la spécification aussi complète que possible du service. Nous décrivons les classes de service de RAP et le dispositif de métrologie associé rendu nécessaire par leur entrée en production. Nous concluons en évoquant la formalisation de l'ensemble dans les contrats de service passés avec les établissements utilisateurs.

Mots clefs

CoS, QoS, MAN, réseau métropolitain, RAP, SLS, SLA

1 Le contexte

1.1 Introduction

Le Réseau Académique Parisien (RAP) raccorde entre eux et vers Renater et Géant 51 établissements répartis sur près de 120 sites parisiens. Pour l'essentiel, RAP assure la connectivité des universités parisiennes et de certains grands établissements, grandes écoles, instituts, centres hospitalo-universitaires, de la plupart des unités parisiennes du CNRS et de l'INSERM, du CROUS et de la Cité Universitaire.

RAP est un réseau majoritairement optique mettant en œuvre le multiplexage de longueurs d'ondes sur son anneau central. Après une phase de déploiement initial concernant essentiellement l'IPv4 unicast jusqu'en 2003 [1], RAP consolide, conformément à ses objectifs du départ, son offre de services complémentaires sur l'ensemble des couches du modèle ISO (VLANs, IPv6, multicast, lissage de trafic, métrologies, supervision, hébergement de listes de diffusion, serveurs vidéo, services de visioconférence ...) dans une démarche de cohérence avec la fourniture des services par le réseau national Renater et le réseau européen Géant. D'une manière générale, la fourniture des services par RAP se fait dans un cadre formalisé, les spécifications de ces services sont

publiées sur le web de RAP (www.rap.prd.fr), permettant de mettre en correspondance la valeur ajoutée apportée par le réseau et les coûts induits auprès des utilisateurs.

1.2 Description et organisation

Sur le plan technique, le centre opérationnel de RAP (le CORAP) est chargé de la mise en œuvre du réseau. Les personnels du CORAP proviennent de la communauté enseignement/recherche ainsi que du prestataire retenu pour opérer le réseau (la société NextiraOne). Un guichet unique assure la réception des signalisations d'incidents et le pilotage des actions de maintenance préventive et corrective impliquant éventuellement des fournisseurs de matériels et de liaisons. Les tâches affectées aux ressources propres comprennent entre autres l'ingénierie du réseau et des services, la communication externe, la relation avec les établissements, la spécification et la mise en place des services avancés ainsi que la gestion courante. Le prestataire NextiraOne assure l'exploitation, la supervision et la maintenance du réseau.

1.3 Architecture technique du réseau

RAP comporte plus de 90 sites raccordés en fibre optique dans Paris ainsi qu'une trentaine de sites accédant au réseau par des liaisons d'opérateurs (« petits sites » raccordés à 2 et 4 Mbit/s). Les sites sont raccordés sur cinq points de présence (POPs) géographiquement distribués dans Paris et hébergés dans des locaux de la communauté enseignement/recherche. Le multiplexage de longueurs d'onde (DWDM) des liaisons optiques inter-POP permet le déploiement des différents services du cœur de RAP. Dans chaque POP, un commutateur-routeur constitue un point de routage pour le trafic IPv4 (plus de 400 réseaux IP routés) et permet l'établissement des VLANs utilisés par certains établissements pour constituer des liaisons virtuelles inter-sites. Dans trois des cinq POPs, des routeurs dédiés assurent le raccordement des réseaux IPv6 entre eux et vers Renater (cf. Figure 1).

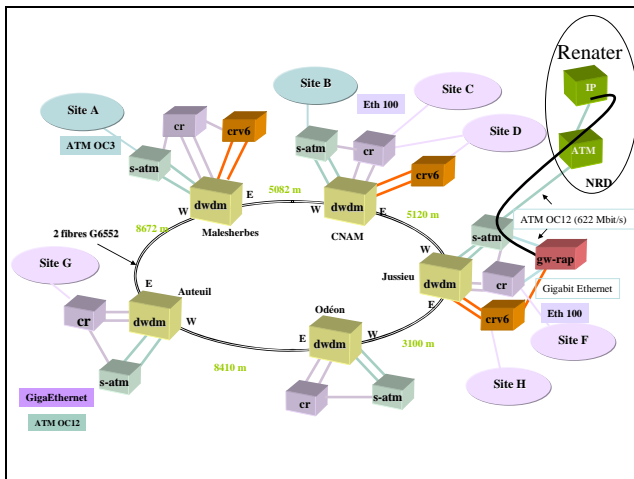


Figure 1 – L'architecture générale de RAP

2 L'expression du besoin de classes de service

La demande de Qualité de Service existe sous différentes formes selon les applications concernées. Elle est implicite si on considère le rôle des réseaux régionaux et métropolitains dans la prolongation des services des réseaux nationaux et internationaux, mais aussi explicite lorsqu'on aborde les questions de transport de la voix et de l'image pour lequel l'utilisateur est sensible à la moindre altération du service. Depuis 2001, plusieurs établissements ont demandé la mise en œuvre de QoS sur RAP pour des applications de téléphonie en particulier. Le Centre Opérationnel (CORAP) lui-même est aussi demandeur pour garantir la fiabilité de ses services de pont de visioconférence. Enfin, l'augmentation permanente de débit sur l'accès vers Renater (40% de la bande passante sur 5 minutes couramment atteints en journée) rend inéluctable la survenue de congestions en ce point pendant des durées de moins en moins courtes.

2.1 Les applications scientifiques

RAP est avant tout un réseau de la communauté scientifique. Comme pour d'autres réseaux régionaux et métropolitains, le premier besoin à satisfaire a donc été de prolonger les dispositifs proposés à des fins scientifiques aux plans européen et national par Géant et Renater [2]. Il s'agit des classes de service dénommées Premium IP et LBE (Less than Best Effort). Sur Géant et Renater, la classe de service Premium IP est strictement prioritaire par rapport aux autres excepté celle utilisée pour le management du réseau, ce qui permet la mise en place de systèmes matériels ou logiciels nécessitant un couplage fort, répartis à l'échelle européenne. Les utilisations sont par exemple l'interférométrie VLBI (Very Long Baseline Interferometry) ou les grilles de calcul. Toujours sur Géant et Renater, la classe de service LBE est strictement moins prioritaire par rapport aux autres, elle sera choisie pour

effectuer des transferts de données très volumineux sans impact pour les autres flux. Les utilisations peuvent concerner des centres de stockage répartis, ou des accès lourds aux bases de données de physique nucléaire.

2.2 L'enseignement à distance

L'enseignement à distance, quand il utilise des moyens de transmission audio et vidéo en temps réel (ou en direct), ne doit pas être perturbé par d'autres flux sur le réseau qui n'ont aucune contrainte forte de temps d'acheminement. Par exemple, il n'est pas admissible que la mise à jour de leur système d'exploitation par les utilisateurs d'un site perturbe le déroulement des cours pour les étudiants de ce site.

Dans certains cas (premières années de médecine par exemple), un cours dont la transmission se passe mal entre deux amphithéâtres devra même être totalement interrompu pour des raisons d'égalité de traitement des candidats au concours de fin d'année.

2.3 Les applications de communication, de surveillance, de gestion technique

A l'occasion du renouvellement des matériels téléphoniques, de plus en plus d'établissements multisites s'orientent tout naturellement vers le transport de la voix par le réseau entre les sites. Sur RAP, une dizaine de sites sont déjà concernés et ce nombre devrait doubler en deux ans. Dans ce contexte, les fournisseurs de solutions de communications ne peuvent pas se contenter de l'hypothèse selon laquelle le surdimensionnement évitera les congestions néfastes à la voix. Ils demandent au réseau quels sont les paramètres garantis quant au délai d'acheminement et à la gigue, en dehors desquels ils n'engageront pas leur responsabilité en cas de dysfonctionnement de leurs produits.

De même, la mise en œuvre d'une solution de vidéosurveillance, de contrôle d'accès aux locaux ou de gestion technique de bâtiment ne pourra se faire qu'avec des caractéristiques de transport adaptées et garanties par le réseau.

Dans ces deux cas, seul l'ordonnement des flux adapté à la criticité de l'application permettra à celle-ci de s'appuyer sur un réseau par lequel transitent toutes sortes d'autres trafics.

2.4 Objectifs techniques associés à l'expression du besoin

Pour les besoins exposés ci-dessus, on évoque des problématiques de temps de transit, de gigue et de taux de perte. Ces éléments constituent donc les objectifs techniques de la mise en œuvre des classes de service, ils sont en général satisfaits par l'ingénierie du réseau (équipements non bloquants, liaisons de capacités adaptées) tant que l'on évite les situations de concurrence

pour l'émission de trafic sur une même interface. Il est important de noter que l'on retrouvera ces objectifs dans la définition des métriques permettant de vérifier la fourniture des classes de service sur RAP (cf. paragraphe 4.5.1).

2.5 Moyens techniques et intérêt des mécanismes de QoS

Si pendant une durée non nulle (quelques fractions de seconde) la somme des débits nécessaires aux applications dépasse la capacité d'émission et de mise en mémoire tampon d'une interface, une ou plusieurs de ces applications ne pourront plus fonctionner de manière optimale pour les utilisateurs. Par exemple, sur une interface 100 Mbit/s déjà occupée à 30%, une diffusion vidéo à 8 Mbit/s en temps réel sera probablement perturbée ou dégradée au démarrage d'un transfert pour un système de sauvegarde répartie accompagné du téléchargement par un utilisateur d'un fichier volumineux.

Il faut donc faire en sorte que les différents flux soient traités d'une manière adaptée. Dans le cas précédent, il faut sans doute réserver la bande passante nécessaire au flux vidéo et ralentir le transfert lié à la sauvegarde. Cette disposition sera probablement sans effet visible pour les utilisateurs. Cette organisation de la priorité des flux entre eux est effectuée par les mécanismes de QoS (Quality of Service) disponibles dans les équipements.

3 Les possibilités de l'infrastructure

L'infrastructure de RAP est composée d'équipements actifs et de liaisons dont nous allons détailler les principales caractéristiques concernant la QoS.

3.1 Les équipements actifs

Nous n'évoquons pas ici les équipements optiques du réseau qui n'interviennent pas du tout dans l'ordonnement des flux.

3.1.1 Les commutateurs du backbone

Les sites optiques de RAP sont raccordés au POP le plus proche sur un commutateur Extreme Networks 6808 présentant des fonctionnalités suivantes :

- Classification des flux dans huit classes de service internes selon les mécanismes suivants :
 - ACL de niveau 2, 3 et 4
 - Tag 802.1P
 - DiffServ (DSCP/IP Precedence)
 - PortL'ordre de prise en compte de ces mécanismes est lui-même paramétrable
- Ordonnement des flux par huit files matérielles sur chaque interface

- Possibilité de fixer des bandes passantes minimales et maximales à chaque file matérielle en sortie
- Pas de lissage ou de limitation par classe de service en entrée dans le commutateur

3.1.2 Le routeur d'accès des sites bas débits

L'ensemble des sites bas débits est concentré par l'opérateur au moyen de PVCs sur une interface ATM OC3. Sur RAP, c'est un routeur Cisco 3725 qui fournit cette interface, et qui est lui-même raccordé au réseau par une interface FastEthernet. Les fonctionnalités fournies par cet équipement sont les suivantes :

- Classification des flux dans de multiples classes de service internes selon les mécanismes suivants:
 - NBAR (classification par type d'application)
 - ACL de niveau 2, 3 et 4
 - Tag 802.1p
 - Diffserv (DSCP/IP Precedence)
 - Champ EXP MPLS (sans objet à ce jour sur RAP)
 - ATM CLP bit
- Ordonnement des flux vers de multiples files matérielles sur chaque interface
- Possibilité de fixer des bandes passantes maximales aux différents flux en entrée et en sortie
- Possibilité de transporter les informations Diffserv dans ATM grâce au standard AAL5

3.1.3 Le routeur d'accès à Renater

L'interconnexion entre RAP et Renater est assurée par un routeur Juniper M7i doté des fonctionnalités suivantes :

- Classification des flux dans quatre classes de service internes selon les mécanismes suivants :
 - ACL de niveau 2, 3 et 4
 - Tag 802.1P
 - DiffServ (DSCP/IP Precedence)
 - Champ EXP MPLS (sans objet à ce jour sur RAP)
- Ordonnement des flux par quatre files matérielles sur chaque interface
- Information supplémentaire PLP (Packet's Loss Priority) permettant de distinguer les flux au sein d'une même file
- Possibilité de fixer des bandes passantes maximales aux différents flux en entrée et en sortie

3.2 Les liaisons

3.2.1 Les liaisons optiques

Sur RAP, les liaisons optiques ne sont pas suffisamment longues pour introduire un délai remarquable dans l'acheminement des trames. Leur nature point à point n'entraîne en temps normal aucune perte d'information. Ces liaisons pourront être négligées dans les problématiques de classes de service.

3.2.2 Les liaisons bas débit

Depuis mai 2005, les accès bas débit de RAP sont des liaisons SDSL fournies par France Télécom (offre TurboDSL). On introduira les paramètres de perte, de temps de transit et de gigue indiqués dans les STAS (Spécifications Techniques d'Accès au Service). On notera que concernant le temps de transit et le taux de perte, l'opérateur n'annonce pas des « garanties » mais des « objectifs de service ». Cette nuance devra être prise en compte lors de la spécification des niveaux de service de RAP pour les sites bas débit.

4 La spécification du service

4.1 Les classes de service

Compte tenu des besoins exprimés, cinq classes de service ont été définies sur RAP, dont voici la liste (de la plus prioritaire à la moins prioritaire) :

4.1.1 La classe Temps Réel

4.1.1.1 Description de Temps Réel

Cette classe est prioritaire sur tout autre trafic utilisateur. Elle est restreinte au périmètre de RAP car elle n'a pas d'équivalent sur Renater. Cette classe est adaptée pour le transport des services et applications critiques entre sites de RAP. Elle est utilisée aujourd'hui pour les interconnexions de PABX, et conviendra aussi pour la vidéosurveillance, le contrôle d'accès ou la retransmission en direct d'amphithéâtre à amphithéâtre.

4.1.1.2 Conditions d'utilisation de Temps Réel

Chaque établissement dispose d'un droit d'utilisation de la classe Temps Réel sur RAP pour un débit global de 2 Mbit/s par site raccordé. L'établissement déclare les adresses IP ou MAC sources et destinations pour ces flux dont la classification est effectuée au moyen d'ACLs (Access Control Lists) installées dans les équipements de RAP.

4.1.2 La classe Premium IP

4.1.2.1 Description de Premium IP

Cette classe est équivalente à la classe Premium IP de Renater et Géant. Sur RAP, elle permet aux utilisateurs d'accéder à une forte priorité sur le réseau pour les machines ou les applications de leur choix.

4.1.2.2 Conditions d'accès à Premium IP

Chaque établissement dispose d'un droit d'utilisation de la classe Premium IP sur RAP pour un débit global de 2 Mbit/s par site raccordé. Le trafic est marqué sur le site de départ au moyen du champ DSCP 46.

Si la classe Premium IP de RAP est compatible avec les classes Premium IP de Renater et de Géant, il faut noter que l'accès à la classe de service Premium IP sur ces réseaux n'est possible que pour les projets scientifiques reconnus. En dehors de RAP, cette classe n'est donc accessible qu'à des conditions assez strictes.

4.1.3 La classe BBE – Better than Best Effort

4.1.3.1 Utilisations de BBE

Cette classe offre une meilleure priorité que celle du trafic Best Effort et permet donc de favoriser certains flux par rapport au trafic général. Elle est compatible avec la classe BBE de Renater (accessible sans formalité particulière) et peut être utilisée avec d'autres sites en France sans nécessairement entrer dans le cadre d'un projet scientifique.

Les utilisations conseillées sont notamment le transport de la voix (par exemple pour des softphones déployés sur des stations de travail non dédiées) et la visioconférence à l'intérieur de RAP ou en passant par Renater.

4.1.4 Conditions d'accès à BBE

Chaque établissement dispose d'un droit d'utilisation de la classe BBE sur RAP pour un débit global de l'ordre de 20% de l'ensemble de ses débits de raccordement à RAP. Le trafic est marqué par le site de départ au moyen du champ DSCP 2.

4.1.5 La classe BE – Best Effort

4.1.5.1 Description de BE

C'est la classe par défaut, à utiliser pour les trafics n'ayant pas de contrainte particulière. On y trouve essentiellement des flux asynchrones (à l'échelle de temps du réseau) régulés par TCP : Web, Mail, FTP ...

4.1.5.2 Conditions d'utilisation de BE

Sur RAP, les sites peuvent émettre et recevoir du trafic Best Effort à hauteur de leur débit nominal de

raccordement au réseau. Pour les sites accédant aux classes de service, l'accès à BE se fait en positionnant le champ DSCP à 0.

4.1.6 La classe LBE – Less than Best Effort

4.1.6.1 Description de LBE

La classe de service LBE est la moins prioritaire de toutes. Le trafic LBE ne pourra s'écouler que si les flux de toutes les autres classes ont été satisfaits. Cela permet de mettre en œuvre des applications très consommatrices en bande passante tout en protégeant le reste du trafic.

4.1.6.2 Conditions d'utilisation de LBE

Chaque établissement dispose d'un droit d'utilisation illimité de la classe LBE sur RAP. Le trafic est marqué par le site de départ au moyen du champ DSCP 8.

Si la classe LBE de RAP est compatible avec les classes LBE de Renater et de Géant, il faut noter que l'accès à la classe de service LBE sur ces réseaux n'est possible que pour les projets scientifiques reconnus. En dehors de RAP, cette classe n'est donc accessible qu'à des conditions assez strictes.

4.2 La protection des classes de service de moindre priorité

4.2.1 Les raisons de la protection

La concrétisation de la QoS sur le réseau est avant tout l'ordonnement des trafics sur les interfaces. Si on s'en tient là, on aboutit à une situation dans laquelle les trafics non prioritaires par rapport à d'autres attendent que ces derniers aient été totalement écoulés pour être acheminés à leur tour. Dans le cas où le volume des trafics prioritaires atteindrait la bande passante nominale du réseau, les autres flux ne seraient jamais servis et finalement perdus après remplissage des tampons qui leur sont affectés sur les équipements. Sur la figure suivante, trois sites sont raccordés au même débit sur réseau. Les sites A et B émettent du trafic à leur débit nominal vers le site C à part égale dans les deux classes de service BBE et BE. Aucun dispositif particulier n'est prévu et comme la somme des débits BBE atteint la capacité d'émission de l'interface vers le site C, aucun trafic BE ne peut passer. Ce taux de perte de 100% ne sera pas acceptable pour le trafic BE (on estime en général qu'il faut une perte inférieure à $1/10^{-5}$ pour une qualité de transport acceptable).

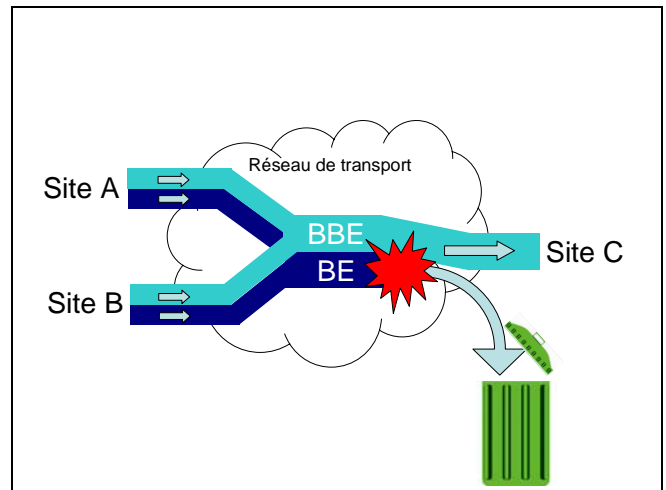


Figure 2 – Exemple de perturbation des trafics BE par des trafics BBE excessifs

4.2.2 Les éléments d'une protection idéale

Pour éviter les situations comme celle décrite ci-dessus, il faut que les volumes de flux dans les différentes classes de service soient limités conformément aux exigences des classes de service de niveau inférieur. Par exemple, en décidant que le trafic BE doit toujours pouvoir bénéficier d'au moins 80% de la bande passante en tout point du réseau. Les dispositifs techniques à mettre en œuvre sont de deux types :

- La limitation de trafic (policing)
Le trafic en dépassement du débit prévu est supprimé.
- Le lissage de trafic (shaping)
Le trafic en dépassement du débit prévu est retardé pour être émis ultérieurement selon le rythme déterminé par le débit prévu.

Par ailleurs, la mise en œuvre de ces dispositifs devrait être effectuée au plus près des sources (dans notre cas les sites A et B) afin de protéger la plus grande partie possible du réseau de transport. Idéalement, le site utilisateur effectuerait un lissage (shaping) de son trafic en émission pour le rendre conforme au contrat de service et le premier équipement du réseau de transport effectuerait une limitation (policing) pour garantir le service aux autres utilisateurs.

4.2.3 La protection mise en œuvre sur RAP

Les équipements de RAP ne permettent aucun lissage de trafic, et ne permettent pas non plus de limitation en entrée sur les interfaces. On ne peut donc que limiter les débits en sortie sur les interfaces entre équipements de RAP et vers les sites. Par exemple, on limite la classe de service BBE à 20% du trafic global sur l'ensemble des interfaces du backbone, et sur les interfaces des sites qui accèdent à cette classe de service.

4.2.4 Les limitations de la protection

Ce dispositif protège globalement les classes de service entre elles mais ne permet pas de garantir la répartition équitable de l'utilisation à l'intérieur d'une même classe. Ainsi, si un site dépasse de manière importante son débit autorisé pour une classe de service faisant l'objet d'une limite sur le réseau, c'est l'ensemble des flux de cette classe de service qui sera écrêté, entraînant ainsi un taux de perte important dans cette classe de service pour les autres utilisateurs. Face à cette éventualité, le déploiement d'une métrologie spécifique aux classes de service (voir paragraphe 5.2) permet de détecter ces anomalies, et de prendre dans un délai maîtrisé les mesures appropriées pour rétablir le service au plus grand nombre.

4.3 La fourniture du service

4.3.1 Les modalités d'accès au service

Les classes de service de RAP sont accessibles aux établissements raccordés dans les conditions décrites ci-dessous et sur demande expresse de leur part. Sans demande particulière, les flux en provenance des sites de l'établissement ainsi que les flux en provenance de Renater vers l'établissement seront traités dans la classe de service BE (Best Effort).

4.3.2 Les niveaux d'accès définis sur RAP

L'accès aux différentes classes de service de RAP se fait par l'inscription du site au niveau d'accès approprié. Le tableau suivant donne les classes accessibles par niveau d'accès :

		Niveaux d'accès aux classes de service			
		Sans accès	Accès de base	Accès Premium IP ⁽¹⁾	Accès Temps Réel ⁽²⁾
Classes de service	LBE	Non	Oui	Oui	-
	BE	Oui	Oui	Oui	-
	BBE	Non	Oui	Oui	-
	Premium IP	Non	Non	Oui	-
	Temps Réel	Non	Non	Non	Oui

(1) Le niveau d'accès « Premium IP » n'est accessible que si l'établissement s'engage à mettre des moyens de contrôle de la classe Premium IP en sortie du site.

(2) Le niveau d'accès « Temps Réel » s'ajoute à un des trois autres niveaux. Il n'est accessible que si l'établissement s'engage à ne générer du trafic dans la classe Temps Réel qu'à partir de systèmes dédiés aux applications de communication et administrés en conséquence.

Figure 3 – Tableau des classes de services fournies en fonction des niveaux d'accès

4.4 La classification des flux et le contrôle du respect des contrats

Compte tenu de la configuration technique et des possibilités des équipements de RAP, la mise en œuvre des classes de service est réalisée de la manière suivante :

Les flux, qu'ils arrivent d'un site RAP ou de Renater, sont classifiés dans une des cinq classes de service par le premier équipement actif qu'ils traversent. Ils se voient alors affectés un tag 802.IP qui sera utilisé par les autres équipements jusqu'à leur destination (site RAP ou sortie vers Renater).

Les flux en provenance de Renater sont classifiés de manière cohérente avec les classes de service utilisées par les sites destinataires. Ainsi, un flux Premium IP sur Renater sera reclassé BE sur RAP si le site destinataire n'accède pas à la classe de service Premium IP sur RAP.

Les flux en provenance de Renater affectés aux classes de service BBE ou Premium IP font l'objet d'une limitation de débit correspondant au contrat de trafic du site destinataire sur RAP.

Il n'est pas possible de limiter les débits en provenance des sites de RAP par classe de service. Une métrologie doit être mise en œuvre pour déceler les dépassements de contrats risquant de dégrader le service aux autres sites dans la même classe de service.

On notera que dans tous les cas, le champ DSCP est laissé intact, ce qui permet à un site de l'utiliser sur RAP pour ses propres besoins, indépendamment des dispositifs décrits dans ce document.

4.5 Preuve du fonctionnement des classes de service

Comme c'est le cas pour les autres services de RAP, la fourniture des classes de service sur le réseau doit pouvoir être vérifiée et mesurée par le Centre Opérationnel comme par les utilisateurs. Dans le cas présent, seule la métrologie active (s'appuyant sur du trafic effectivement émis) permet de vérifier les performances du réseau pour les différentes classes de service. De plus, le trafic de métrologie active doit être traité de manière équivalente au trafic de production. Sur RAP, cette dernière condition exclu tous les outils de validation du réseau à base d'ICMP, que les commutateurs BlackDiamond acheminent de manière différente des autres trafics.

Il est donc nécessaire de définir les métriques utilisées dans la caractérisation du service pour ensuite spécifier un outil de base permettant les mesures appropriées sur le réseau.

4.5.1 Les métriques retenues

Les éléments retenus pour décrire les performances du réseau à l'intérieur des classes de service sont les suivants :

- Le temps aller-retour ou RTT (Round Trip Time)

La mesure du temps de transit (aller simple) nécessite la mise en œuvre de mécanismes coûteux de synchronisation des horloges aux deux extrémités de la mesure, ce qui n'est pas le cas de la mesure du temps aller-retour. Sur RAP, on utilise la mesure du temps aller-retour pour vérifier que l'aller simple n'excède pas les valeurs prévues (le temps de l'aller simple ne peut pas être supérieur à celui de l'aller retour).

- La gigue
- Le taux de perte

4.5.2 L'outil de métrologie active

Les critères qui ont guidé le choix de l'outil de métrologie active ont été les suivants :

- L'outil doit fournir les métriques définies ci-dessus.
- Les mesures se feront dans chaque classe de service, l'outil doit générer les champs DSCP correspondants.
- Les mesures pourront être faites entre un site et un point de référence sur RAP, ou entre deux sites. Les sites concernés peuvent se situer à l'extérieur de RAP.
- Les mesures seront enregistrées pour permettre l'examen a posteriori du comportement du réseau.
- L'utilisation sera simple et le logiciel utilisé sera libre de droits.

5 Le déploiement des classes de service sur RAP

Nous allons décrire ici les trois principaux volets de la mise en œuvre des classes de service sur RAP, c'est-à-dire :

- La configuration des mécanismes de QoS sur les équipements actifs
- Le dispositif de métrologie passive de l'utilisation des classes de service par les sites, qui permet de vérifier que le trafic reste conforme avec les modalités d'accès définies sur RAP
- Le dispositif de métrologie active, qui permet de vérifier que les paramètres de fonctionnement du réseau sont conformes aux spécifications des classes de service

5.1 La configuration du réseau

Nous ne décrivons ici que la mise en œuvre de la QoS sur les équipements Extreme Networks BlackDiamond 6808. Des démarches similaires ont été menées pour les équipements Juniper et Cisco.

5.1.1 La validation sur maquette

En février et mars 2005, une maquette a été constituée pour tester et valider les mécanismes de QoS prévus. La figure suivante présente le schéma général de cette maquette :

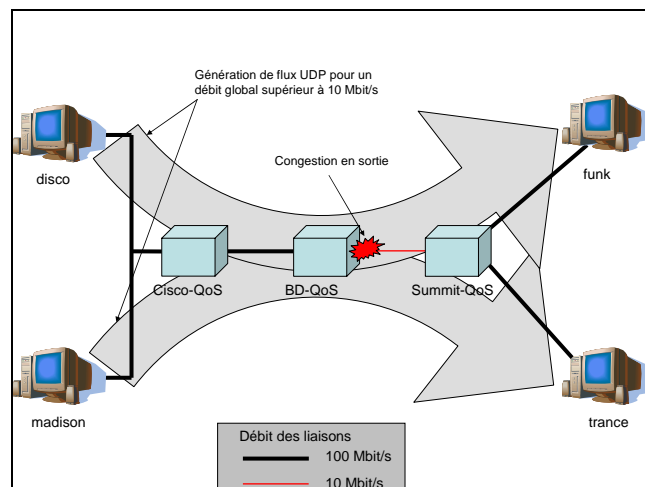


Figure 4 – Schéma de la maquette de validation de la QoS sur BD 6808

La maquette a mis en œuvre 3 commutateurs et 4 stations :

Cisco-QoS : commutateur Cisco 3550

BD-QoS : commutateur Extreme Networks BlackDiamond 6808

Summit-QoS : commutateur Extreme Networks Summit 48i

madison et funk : PC linux Fedora core III

disco : PC Linux Mandrake 10

trance : Station Sparc Solaris 8

Le principe des tests était de générer avec l'outil IPERF deux flux pour un débit total supérieur à la capacité du lien vers Summit-QoS (10 Mbit/s) et de déclencher ainsi les mécanismes d'ordonnement de BD-QoS. IPERF donnait un premier résultat sur les débits effectifs alloués par le réseau. D'autres points ont pu être vérifiés à l'aide des commandes internes des commutateurs (notamment qosmonitor sur le BlackDiamond) et de l'analyseur réseau EtherReal.

Les éléments suivants ont été validés :

Mécanismes de QoS :

Classification par port, par ACL de niveau 2 et 3 et par le champ DSCP

Ordonnement attendu pour les flux traités aux niveaux 2 et 3 par le commutateur

Fonctionnement des limitations de bande passante pour permettre le passage des classes de service de rangs inférieurs

Métrologie passive des classes de service :

Présence des informations de QoS dans les informations Sflow émises par le BD-QoS

Métriologie active :

Positionnement correct du champ DSCP par IPERF

La maquette ayant permis de valider l'ensemble des mécanismes prévus sur les équipements Extreme Networks, le déploiement sur le réseau de production a pu être envisagé.

5.1.2 Le déploiement sur le réseau de production

A partir de mars 2005, les commutateurs du backbone de RAP ont été configurés pour prendre en compte les mécanismes de QoS. Cependant, des problèmes affectant la connectivité de certains sites sont survenus sur le réseau de production, pour des configurations identiques à celles validées sur la maquette. Il a finalement été nécessaire de remplacer les cartes d'interfaces de raccordement des sites pour des versions plus récentes.

Le service déployé est stable depuis juin 2005.

5.2 La métriologie des classes de service

Comme nous l'avons vu aux paragraphes 3.1.1 et 4.4, les commutateurs BlackDiamond de RAP ne permettent pas de limiter le trafic par classe de service en entrée sur ses interfaces. Cela introduit le risque qu'un site en dépassement de son contrat provoque une perturbation pour l'ensemble des utilisateurs de la classe de service concernée en raison de la limitation globale associée à la classe de service sur le backbone. L'enjeu principal de la métriologie des classes de service est de détecter une telle situation et d'y remédier dans le cadre de la garantie de temps de rétablissement (4 heures) associée aux classes de service.

Les outils NFSen (Netflow Sensor - <http://nfsen.sourceforge.net>) et Sflowtools (<http://www.inmon.com/technology/sflowTools.php>) ont été déployés par le CORAP pour mesurer les trafics par classe de service sur la base des flows transmis par les différents équipements du réseau. On dispose ainsi, pour chaque site des graphes instantanés et consolidés dans le temps indiquant les débits émis par le site dans chaque classe de service. La figure ci-dessous présente un exemple de configuration pour la métriologie de classe de service d'un site dans NFSen. Afin d'automatiser les procédures, il est prévu que la configuration de NFSen soit générée automatiquement à partir du système d'information de RAP (SIRAP) [3] en fonction des services fournis au site.

Profile: A06_BBE		Le nom du profil correspond au code du site suivi de la classe de service
Description:	Trafic émis par le site A06 - ENSAM dans la classe de service BBE	
Type:	Continuous	On peut créer des profils consolidant les mesures en provenance de plusieurs points du réseau
Start:	2005-09-20-11-05	
End:	2005-09-20-11-05	
Last Update:	2005-09-20-11-00	Le site qui fait l'objet de la mesure est désigné par ses Classes IP
Sources:	pop_jussieu pop_eulieuil pop_odeon	
Filter:	(SRC NET 192.93.173.0/24 OR SRC NET 193.50.216.0/24 OR SRC NET 193.50.217.0/24) AND TOS 8	Dans le cas de BBE, c'est le Champ DSCP qui permet la classification
Size:	0 B	
Max. Size:	unlimited	
Expire:	never	
Status:	OK	

Figure 5 – Exemple de configuration, d'un profil dans NFSen

Voici un exemple de courbe produite par NFSen (et intégrée dans le portail utilisateur de RAP) montrant le trafic dans la classe Temps Réel pour un site.

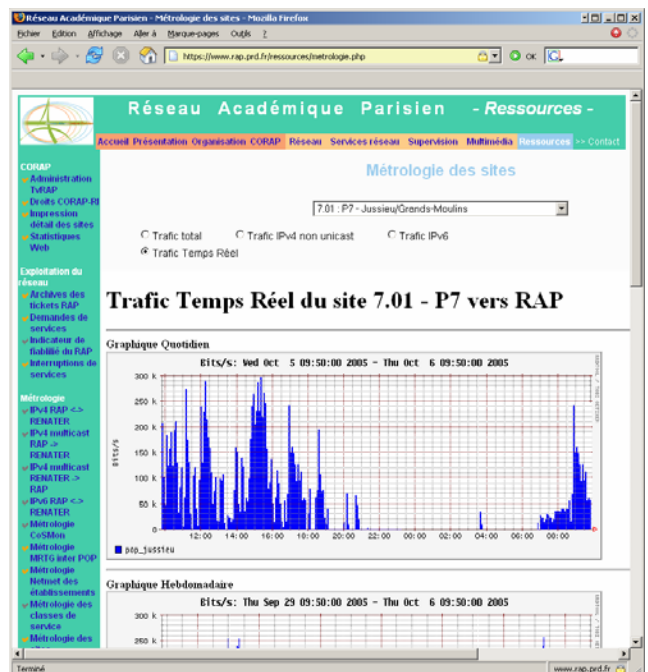


Figure 6 – Intégration de la courbe NFSen dans la métriologie des sites du portail RAP

5.3 La métriologie active

Nous avons vu au paragraphe 4.5 les caractéristiques souhaitées pour l'outil d'évaluation du fonctionnement des classes de service (métriques et caractéristiques de l'outil).

En fonction de ces critères, le CORAP a réalisé une intégration des logiciels IPERF et RRDTools en un package dénommé COSMON. IPERF ne permettant pas la

mesure du temps de transit, on considèrera dans un premier temps que celui-ci est du même ordre de grandeur que la gigue (une version ultérieure de COSMON utilisera l'outil UDPMON en remplacement d'IPERF pour mesurer le RTT)

L'idée est de générer régulièrement (toutes les 5 minutes) entre deux points du réseau un transfert IPERF en UDP dans la classe de service souhaitée (comme pour le trafic de production, la classification est assurée par le champ DSCP sauf pour la classe Temps Réel pour laquelle ce sont les adresses IP/Mac sources ou destinations qui déterminent la classe de service). A l'issue du transfert IPERF, la gigue et le taux de perte sont enregistrés dans une base de données RRD de laquelle on tire par la suite les graphes instantanés et consolidés présentant l'évolution de ces mesures au cours du temps.

La package COSMON et sa documentation sont accessibles depuis le site web de RAP (<http://www.rap.prd.fr/services/metrologieActive.php>).

La figure ci-dessous est générée par l'outil COSMON. Elle présente la Gigue pour la classe de service BE entre une machine du CORAP et une machine située sur le backbone de RAP.

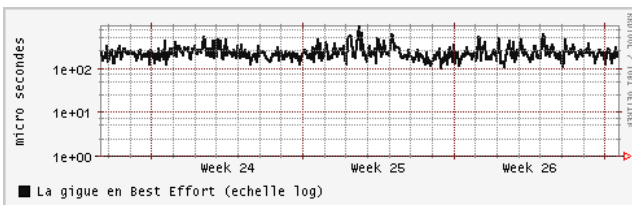


Figure 7 – Gigue sur un mois dans la classe de service BE pour le CORAP

6 La formalisation du service à l'intérieur des SLS de RAP

Un des objectifs de la mise en place des classes de service sur RAP était de garantir des paramètres de trafics adaptés aux différentes applications sur le réseau. L'aboutissement de cette démarche a logiquement été d'intégrer ces paramètres dans les spécifications de niveau de service (ou SLS – Service Level Specifications) portées en annexe des conventions de service entre les établissements utilisateurs et RAP.

6.1 Trafics concernés

Les objectifs et engagements suivants concernent les flux IPv4 unicast pour les protocoles UDP, TCP et GRE (encapsulation IP dans IP utilisée pour construire des tunnels) et les flux IPv4 multicast pour le protocole UDP. Ces garanties s'appliquent également aux flux traités uniquement au niveau 2 du modèle OSI par RAP.

6.2 Objectifs et engagements de performances par classe de service

Les objectifs et engagements suivants concernent le trafic à partir de son émission par l'équipement de raccordement du site jusqu'à son arrivée sur l'équipement du site destinataire (ou sur le premier équipement hors de RAP pour du trafic vers l'extérieur de RAP). Si le site destinataire est raccordé à un débit inférieur à 100 Mbit/s, il convient d'ajouter 5 ms au temps de transit, 10^{-7} au taux de perte et 1 ms à la gigue.

6.2.1 Objectifs pour les classes de service LBE, BE et BBE

RAP est conçu pour fournir au trafic de chaque site une capacité de transit sur le backbone correspondant au débit d'accès pour les classes LBE, BE et à 20% de celui-ci pour BBE. Le tableau suivant indique les objectifs de performances de RAP concernant les paramètres de trafic de la classe BE (compte tenu de l'implémentation des classes de service, les paramètres seront moins favorables pour LBE et plus favorables pour BBE).

	Raccordement de l'émetteur > 100 Mbit/s	Raccordement de l'émetteur < 100 Mbit/s
Temps de transit maximum (ms)	6	11
Temps de transit moyen (ms)	2	7
Taux de perte	$<10^{-7}$	10^{-7}
Gigue (ms)	3	4

6.2.2 Engagements pour les classes de service Premium IP et Temps Réel

Les classes de service Premium IP et Temps Réel sont fournies aux sites pour des trajets définis sur RAP. Elles font l'objet d'une pré-allocation des ressources du backbone et les liaisons vers les sites qui permet au CORAP de s'engager sur le respect des paramètres de fonctionnement suivants.

	Raccordement émetteur > 100 Mbit/s	Raccordement émetteur < 100 Mbit/s
Temps de transit maximum (ms)	2	7
Temps de transit moyen (ms)	2	7
Taux de perte	$<10^{-9}$	$<10^{-9}$
Gigue (ms)	< 2	2

6.3 Garanties de temps de rétablissement

En dehors d'un problème de liaison ou d'un dysfonctionnement sur les équipements du réseau, le risque le plus important de perturbation des classes de service provient de l'émission abusive de trafic de la part d'un site dans une classe de service. Ceci amènerait l'ensemble des flux de la classe de service à la limite maximale fixée sur le backbone et il s'en suivrait automatiquement un taux de perte élevé pour tous les utilisateurs.

A défaut de pouvoir éviter totalement le problème, le prestataire opérant RAP est habilité à utiliser la métrologie par classes de service pour identifier le site responsable du problème, et lui supprimer à titre conservatoire l'accès aux classes de service. Cette mesure permet non seulement de rétablir le service normal pour les autres utilisateurs, mais aussi de fournir un service moins dégradé au site en question. Cette action doit être menée dans une durée contractuelle de 4 heures.

Dans ce cas, il va de soit que le CORAP assistera le site dans la recherche rapide de l'origine du problème et dans sa résolution afin de rétablir l'accès aux classes de service.

7 Bilan

La démarche de mise en œuvre des classes de service sur RAP a duré un peu plus de neuf mois, occupant principalement un ingénieur à 30% de son temps. Les tâches réalisées se répartissent équitablement en trois catégories :

1/ L'étude technique et la validation de la mise en œuvre sur les équipements (Extreme Networks, Juniper, Cisco)

2/ La conception et l'intégration de systèmes de métrologie accompagnant le déploiement (NFSen, Sflowtools, IPERF, RRDTOOLS). Les correspondants techniques des sites utilisateurs accèdent directement à cette métrologie depuis le portail de services de RAP (<http://www.rap.prd.fr/ressources>).

3/ L'industrialisation (spécifications, procédures, demandes de services au prestataire) du service sur le réseau de production et l'accompagnement des premiers utilisateurs du service

Comme c'est parfois le cas pour d'autres services avancés (multicast par exemple), le déploiement des classes de service sur le réseau de production a mis en évidence d'importants problèmes sur les équipements. Dans le cas de RAP, si on a pu revenir au niveau de fiabilité normal du réseau en remplaçant l'ensemble des cartes d'interface de raccordement des sites, cela a mis en retard de près de deux mois la fourniture du service en production.

La mise en production a finalement eu lieu en juin 2005 et le service est déjà utilisé par plus d'une dizaine de sites dont la moitié pour des interconnexions de PABX. Par

ailleurs, les services de pont de visioconférence de RAP ont aussi été configurés pour tirer profit du dispositif.

Depuis mi-2005, les conventions de service entre RAP et les nouveaux établissements utilisateurs intègrent les classes de service, l'extension à toutes les conventions aura lieu dès 2006.

Bibliographie

[1] Jean-Paul GAUTIER – Réseau Académique Parisien, RAP – Un réseau optique. Dans *Actes du congrès JRES2001*, Lyon, Décembre 2001

[2] Franck SIMON – GIP RENATER, Implémentation des classes de services dans RENATER-3. . Dans *Actes du congrès JRES2003*, pages 125-131, Lille, Novembre 2003

[3] Laurent GYDE et Nicolas MENECEUR – Réseau Académique Parisien, SIRAP : Le système d'information de RAP. Dans *Actes du congrès JRES2003*, pages 385-396, Lille, Novembre 2003