INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team Rap

Réseaux, Algorithmes et Probabilités

Rocquencourt

THEME 1B

Table of contents

1. Team

Head of project team
Philippe Robert [DR]

Vice-head of project team
Fabrice Guillemin [France Télécom R&D, Lannion]

Administrative assistant
Virginie Collette [TR]

Staff member
Christine Fricker [CR]

Research scientist (partner)
Jacqueline Boyer [France Télécom R&D, Lannion]
Danielle Tibi [Université Paris VII]

Post-doctoral fellow
Nelson Antunes [Université de Faro]
Abdel Ben Tahar [Université de Casablanca]

Ph. D. Student
Youssef Azzana [Université Paris VI]
Nadia Benazzouna [France Telecom R&D Lannion]
Hanène Mohamed [Université Paris VI]

2. Overall Objectives

L'avant-projet "*Réseaux, Algorithmes et Probabilités*" (RAP) vise à formaliser et renforcer une collaboration engagée depuis plusieurs années entre des ingénieurs de France Telecom R&D à Lannion et une équipe de l'INRIA-Rocquencourt. L'objectif est d'engager une collaboration *continue* entre les deux équipes.

La démarche générale de cette proposition de projet consiste à étudier des sujets très bien délimités sur une période de l'ordre de quatre ans. Les centres d'intérêt actuels de l'avant-projet RAP sont

1. Le contrôle d'admission à l'entrée d'un réseau IP. Voir la section 3.3;
2. La métrologie. Voir la section 3.2;
3. L'allocation de bande passante à l'intérieur du réseau (réservation et équité). Voir la section 3.4.

Sur le plan fondamental, les méthodes de renormalisation des processus de Markov issues de la physique statistique sont au cœur des recherches développées par l'avant-projet pour l'étude des réseaux. Voir la section 3.1.

L'articulation de ces activités est, de façon succincte, la suivante: les mesures sur le trafic du réseau Internet (voir la section métrologie) sont à la base des réflexions algorithmiques menées pour le contrôle d'admission et l'allocation de bande passante. Une campagne de mesure est actuellement menée pour évaluer les impacts respectifs du trafic "lourd" (les éléphants) et du trafic "léger" sur le partage du trafic pour concevoir un algorithme de contrôle d'admission des éléphants (voir le travail préliminaire [9]). L'étude qualitative se fait avec des techniques de renormalisation, en supposant qu'un paramètre tend vers une valeur critique, par exemple, le trafic ou la taille du réseau deviennent très grands, ou le taux de perte très petit.

3. Scientific Foundations

3.1. Méthodes de renormalisation

Key words: *Renormalisation des processus de Markov, limites fluides, théorèmes central limite fonctionnels, physique statistique*.

Les trafics qui traversent les réseaux de communication sont d'une extrême hétérogénéité : données, voix, vidéo, etc. Les requêtes en bande passante sont par conséquent hautement variables, de l'ordre de quelques kbits/sec à plusieurs dizaines de Mbits/s. L'impact de cette extrême variabilité est, pour l'instant, assez peu analysé sur le *comportement global* d'un réseau.

Jusqu'au début des années 90, il était couramment admis que ce type de situation n'était pas en rupture avec le cadre des réseaux classiques où les requêtes des trafics à un nœud donné sont statistiquement proches. Il était toutefois bien connu que l'état d'équilibre de ces réseaux est beaucoup plus difficile à caractériser que celui des réseaux classiques. Les études de Rybko et Stolyar (1992), de Lu et Kumar (1991) et de Bramson (1994) ont, par la suite, complètement changé ce point de vue. Elles montrent que l'hétérogénéité statistique seule peut déstabiliser un réseau : même si, pour chaque nœud du réseau, la charge *moyenne* de travail qui arrive est strictement plus petite que sa capacité, le réseau peut osciller de telle sorte que le nombre total de requêtes dans le réseau diverge. Pour ces contre-exemples, chaque nœud du réseau se vide une infinité de fois mais globalement le réseau diverge. Cette situation est impossible dans les réseaux classiques. Ces réseaux avec des trafics hétérogènes sont regroupés sous l'appellation *réseaux multi-classe*. Les processus de Markov associés à ces réseaux multi-classe sont très délicats à étudier, même en ce qui concerne le comportement macroscopique.

Des techniques de renormalisation sont actuellement utilisées pour étudier le comportement au premier ordre de tels réseaux. Si l'espace d'états du réseau est donné par un ensemble $\mathcal{S}$ muni d'une norme $\|\cdot\|$ (typiquement $\mathbb{R}_+^d$) et, si pour $t \geq 0$, $X(x,t)$ décrit l'état de celui-ci à l'instant t quand son état initial vaut x, le processus renormalisé $\overline{X}$ associé est donné par

$$\overline{X}(x,t) = \frac{X(x, \|x\|t))}{\|x\|}.$$

Le temps est accéléré proportionnellement à la taille de l'état initial, la variable spatiale étant renormalisée avec l'inverse de cette taille. Remarquer que l'état initial du processus $(\overline{X}(x,t))$ est de norme 1. Les idées de renormalisation sont anciennes, notamment en physique statistique, elles permettent d'étudier les comportements transitoires de systèmes de particules. Dans le domaine des réseaux, elles ont émergé de façon explicite récemment. Les discontinuités naturelles de la dynamique des réseaux (dues au fait qu'une file d'attente vide ne traite plus de requêtes) sont l'élément distinctif du cadre de la physique statistique. Elles posent des problèmes nouveaux tout à fait intéressants.

Le comportement macroscopique de l'état du réseau s'étudie alors en faisant tendre la norme de l'état initial, $\|x\|$, vers l'infini. Une *limite fluide* $(L(t))$ est une des valeurs d'adhérence de $\overline{X}$ quand la norme de l'état initial x tend vers l'infini. Par exemple, si $X(x,t)$ est une marche aléatoire dans $\mathbb{R}$ dont la moyenne des accroissements vaut δ, la seule limite fluide positive possible est donnée par la fonction $t \to 1 + \delta t$. La renormalisation a gommé toutes les fluctuations pour ne garder que la dérive moyenne. La marche aléatoire $(X(x,t))$ peut être vue comme une perturbation stochastique de la fonction $t \to 1 + \delta t$. Pour une large classe de réseaux, ce point de vue peut être généralisé : l'état renormalisé du réseau converge vers la solution d'une équation différentielle *déterministe* ordinaire. L'état du réseau peut être vu comme une perturbation stochastique de cette solution. Dans le cadre de processus diffusifs, ces perturbations ont été très étudiées, voir par exemple Khasminski (1960) et Freidlin et Wentzell (1979). Dans le cadre des réseaux, Dai (1995) a formalisé le cadre des équations différentielles déterministes qui pouvaient être obtenues. De nombreux travaux consacrés à l'étude des réseaux multi-classe ont suivi. Pour résumer, l'étude des limites fluides a principalement deux avantages :

1. Décrire le comportement macroscopique du réseau i.e. le système dynamique qui décrit le réseau au premier ordre;
2. Donner un critère de stabilité du réseau. En effet, s'il est possible de montrer que toutes les limites fluides sont nulles à partir d'un certain rang, un résultat de Filonov/Rybko et Stolyar montre que le réseau "normal" (i.e. non renormalisé) atteint un état d'équilibre.

3.1.1. Les insuffisances du cadre actuel

Les techniques de limites fluides se sont généralisées au cours des dix dernières années et ont contribué à une meilleure compréhension de la dynamique des réseaux avec des trafics hétérogènes. C'est actuellement un outil incontournable dans ce type d'étude. Il n'en reste pas moins que la connaissance que nous avons actuellement des réseaux multi-classe est encore très parcellaire, de nombreux aspects importants sont encore obscurs : par exemple, le comportement d'un réseau multi-classe sans trafic prioritaire avec seulement deux nœuds n'est actuellement pas connu, même au niveau macroscopique (fluide). Ceci est dû principalement aux raisons suivantes :

1. *L'aléatoire résiduel*. Sur les questions de renormalisation, l'idée qui prévaut actuellement est la suivante : L'état d'un réseau est une perturbation stochastique d'une fonction déterministe. Autrement dit, la résolution d'une équation différentielle déterministe permet d'obtenir le comportement macroscopique du réseau (quitte à éliminer des "fausses solutions" au passage). Si cette approche est effective dans de nombreux cas de réseaux multi-classe, en particulier les réseaux avec des priorités, elle ne couvre pas la majeure partie des applications. En effet, si la renormalisation gomme toutes les fluctuations à la limite, *elle ne supprime pas toutes les composantes aléatoires*. Certaines des composantes aléatoires de ces réseaux ne font pas partie de la partie diffusive et donc restent après le passage à la limite. C'est un problème important pour l'étude du comportement des réseaux. Il est généralement méconnu et peut être mal interprété au niveau des limites fluides en terme de solutions déterministes multiples, alors qu'il n'y a qu'une seule limite fluide, mais aléatoire.
2. *La dimension infinie*. Pour représenter l'état d'un nœud servi par la discipline FIFO d'un réseau où arrivent des trafics de différentes classes, il est nécessaire de connaître la classe $c \in \mathcal{C}$ de la requête à la première place dans la file d'attente, de même pour la deuxième place, etc. L'état du nœud est donc représenté par une chaîne de caractères (c_i) où c_i est la classe du i-ième client dans la file d'attente. L'espace d'états est celui des suites finies de caractères à valeurs dans un espace fini $\mathcal{C}$. Il est bien sûr dénombrable mais inclus dans un espace de dimension infinie $\mathcal{C}^{\mathbb{N}}$. Il n'est donc plus question de résoudre, de façon ultime, une équation différentielle dans un espace $\mathbb{R}^d$. En fait, même le cadre des équations différentielles en dimension infinie n'est pas le cadre naturel. L'élément important pour ces réseaux est que l'évolution du nombre des requêtes de chaque classe ne se décrit pas facilement. Il faut plutôt se tourner vers l'évolution des schémas des chaînes de caractères décrivant le nœud. Ces systèmes sont très délicats à étudier, nombre de notions sont encore à définir pour poser correctement les bases d'une définition correcte de la renormalisation de ces réseaux. Il y a très peu de travaux dans ce domaine (en dehors de ceux de Bramson). Voir les travaux de Gajrat *et al.* [15] sur l'évolution de certaines chaînes de caractères qui étendent ceux de Dynkin et Maljutov dans le cas des marches aléatoires sur le groupe libre. Les applications de ces travaux aux réseaux multi-classe sont cependant limitées : les réseaux correspondants ont un seul nœud et la dynamique ne dépend que d'un nombre borné de caractères au début de la chaîne. Dans un cadre spécifique, Dantzer et Robert ont introduit plusieurs notions qui semblent pouvoir contribuer aux fondements d'une étude systématique de ces réseaux : les notions d'état initial régulier et lisse notamment. En tout état de cause, ces résultats partiels doivent être poursuivis pour dégager une méthode générale de traitement des processus de Markov à valeurs dans les chaînes de caractères.

Les deux aspects mentionnés ci-dessus nous semblent très importants pour comprendre les phénomènes spécifiques aux réseaux de communication traversés par des trafics hétérogènes. La relation entre l'instabilité

du réseau et la divergence des limites fluides associées est un autre point important encore obscur de ces réseaux multi-classe. Il y a en effet le résultat de Filonov, Rybko et Stolyar qui établit une relation entre la stabilité du réseau et le fait que toutes ses limites fluides reviennent à 0 et y restent. Le fait que la divergence des limites fluides entraîne l'instabilité du réseau n'a pas encore été démontré, il y a seulement quelques résultats très partiels dans ce domaine. Cette question qui est liée aux aspects vus dans le point (1) ne fait pas l'objet d'investigations pour le moment.

3.2. Métrologie

Key words: *Traces des flots TCP, Mesures passives.*

Le projet RNRT "Métropolis" qui a commencé le 1er septembre 2001 regroupe le département réseau du LIP6, l'Institut Eurecom, France Telecom R&D, le Groupe des Écoles des Télécommunications (GET), le LAAS, RENATER et l'INRIA. Pendant la durée de ce projet, des expériences seront menées sur le trafic IP sur plusieurs sections du réseau RENATER entre les centres de Lannion, Paris, Toulouse et Nice. Il faut noter que les mesures disponibles actuellement sur le réseau n'ont pas le degré de précision de celles qui seront effectuées dans Métropolis.

Des mesures très précises flot par flot seront effectuées pour ensuite pouvoir discriminer le trafic global : trafic "lourd" (les *éléphants*) ou "léger" (les *souris*), détecter les engorgements locaux, donner les statistiques des processus de perte (notamment caractériser la distribution de la taille des groupes de paquets perdus en cas de congestion), évaluer l'impact de la phase de *slowstart*, etc... De plus, la validation de résultats obtenus dans [14] peut être envisagée en utilisant cette campagne intensive de mesures sur le réseau. C'est une partie très importante de la démarche engagée par RAP. Il s'agit principalement de

1. dégager des résultats *constructifs* sur la description du trafic observé dans un réseau IP. Les résultats actuels des travaux dans le domaine de la métrologie sont essentiellement négatifs: sur le caractère non poissonnien du trafic Internet, que les corrélations ne décroissent pas de façon exponentielle, ni polynomiale, etc...
2. valider les comportements *qualitatifs* prédits par les résultats des parties 3.3.

projet RNRT Métropolis est de dégager une description mathématique aussi simple que possible de certains types de trafic qui permette une analyse quantitative. L'objectif est, dans un premier temps, d'avoir une validation qualitative des comportements étudiés.

3.3. Contrôle d'admission

Key words: *Allocation de bande passante, Algorithmes MaxMin, Équité.*

Le cadre de cette étude est celle d'un routeur à l'entrée du réseau où l'opérateur doit décider de l'acceptation ou non de demandes de connexions caractérisées, éventuellement, par des paramètres de trafic. Ce cadre générique est valable aussi bien dans les architectures Intserv, MPLS ou même DiffServ si une déclaration explicite est effectuée d'une manière ou d'une autre. Il s'agit d'accepter suffisamment de connexions pour maximiser l'utilisation des infrastructures et en même temps contrôler la charge du réseau de telle sorte que les différents niveaux de garantie de service demandés soient satisfaits. À chaque requête, il s'agit de décider si l'occupation du réseau permet d'accepter le niveau de qualité de service demandé par la requête : bande passante, taux de perte, etc... (On se place bien sûr dans le cadre où les mécanismes de réservation de bande passante sont utilisés). L'algorithme d'acceptation au niveau du routeur doit être simple, ce qui se traduit dans ce cas par un minimum de calcul : typiquement une addition et une comparaison avec une valeur critique.

Ce domaine est important pour la gestion d'un réseau, il fait actuellement l'objet de nombreuses investigations. Dans cette perspective, la notion de gestionnaire de bande passante, *Bandwidth Broker*[1]" (BB) a

[1]L'appellation *Bandwidth Broker* est aussi quelquefois utilisée dans le cadre très différent de la négociation de bande passante au sens financier du terme (options,...).

été récemment introduite par Jacobson. Il s'agit, au niveau du domaine d'un ISP dans un contexte Diffserv, d'implanter un agent capable de :

- authentifier la demande d'une requête sur les routeurs d'entrée;
- vérifier que le niveau de service requis est compatible avec l'état du réseau de l'ISP (Contrôle d'admission);
- configurer les routeurs ({Egress,Ingress}-routers) sur la frontière avec les autres ISP de telle sorte que les trafics reçus et envoyés entre ISP soient conformes aux accords passés entre les différents opérateurs (fonctionnement bilatéral).

Le contrôle d'admission est bien entendu l'élément crucial de ce type d'agent.

3.3.1. Contrôle d'admission du trafic élastique

Le cadre est celui d'un lien entre un réseau d'accès (type ADSL par exemple) et Internet. Sur ce lien, plusieurs flots avec des caractéristiques variables sont multiplexés. On s'intéresse aux flots du trafic élastique (le trafic best effort actuel). Ces flots coexistent (via TCP) en étant contraints par la capacité réduite du lien d'accès. Certains flots sont longs (comme les transferts "peer to peer"), ce sont les *éléphants*, le mécanisme de contrôle de la congestion de TCP fait qu'ils adaptent leur débit de transmission à l'état du lien. Cette adaptation revient, en première approximation, à un partage égalitaire de la bande passante disponible. Les autres flots sont courts, moins de vingt paquets, ce sont les *souris* qui, au niveau TCP, ne dépassent pas l'étape de "slow start". Ces flots ne s'adaptent pas à l'état du réseau en raison du petit nombre de paquets transmis. Globalement, le trafic peut être décrit de la façon suivante : les souris dévorent une partie de la bande passante, la partie résiduelle est partagée équitablement entre les éléphants. Cette description macroscopique de l'état d'un lien est celle proposée initialement par l'équipe de J. Roberts [16]. Elle nous semble particulièrement adaptée pour ce type d'étude.

Les mesures menées sur le réseau montrent que la statistique de la taille des transferts des éléphants ont une queue de distribution lourde, i.e. la probabilité que la quantité transférée dépasse la valeur x ne décroît pas exponentiellement mais plutôt de façon polynomiale. Cette caractéristique est très importante, en effet si N éléphants occupent le lien, chacun d'eux reçoit un débit proportionnel à $1/N$. Si la quantité N est assez grande, cela implique que le temps de transfert devient de plus en plus long. Les mécanismes de contrôle de TCP déclenchent un arrêt de la connexion lorsque ce temps excède un certain seuil. Ce phénomène peut se représenter par le fait que chacune des connexions a un temps d'impatience au-delà duquel elle s'interrompt Dans ce contexte, l'idée de base, voir aussi Roberts *et al.* [13], est de limiter au maximum le nombre de connexions interrompues de la sorte par le biais d'un contrôle d'admission. Il s'agit de rejeter les connexions à l'entrée du réseau de façon à réduire au maximum l'utilisation de la bande passante par des connexions qui vont finalement être arrêtées.

Pour que des algorithmes simples et efficaces puissent être conçus dans ce domaine, il est crucial, dans un premier temps, d'étudier l'interaction entre le partage égalitaire et les phénomènes d'impatience. Actuellement, en dehors des simulations, les travaux sont très rares dans ce domaine. Dans une deuxième étape, il convient d'intégrer le trafic des souris qui introduit la variabilité de la capacité de la bande passante offerte aux éléphants.

3.3.2. Contrôle d'admission du trafic prioritaire

Ces questions ont déjà été étudiées en détail dans les réseaux ATM pour le trafic VBR. Le cadre habituel est celui d'un nœud où arrive une superposition de plusieurs types de trafic (définis chacun par leur débit crête et la taille des rafales). Il s'agit de tester si l'acceptation d'un nouveau flot maintient la probabilité de perte d'un paquet en-dessous d'une valeur critique. En théorie, il est possible de calculer dans cette configuration la probabilité que des paquets soient perdus en résolvant une équation de point fixe qui n'est pas triviale. Cette solution n'est pas acceptable car elle ne permet pas de traiter en temps réel les multiples sorties et arrivées au nœud, il faudrait dans ce cas recalculer le point fixe à chaque fois.

Les travaux de Guérin et Elwalid ont permis de dégager une solution acceptable algorithmiquement. À chaque type de flot est associé un nombre, appelé bande passante effective, calculé une fois pour toutes et le nœud maintient un nombre W représentant son occupation. Quand une connexion s'achève, la bande effective correspondante est retranchée de W. À l'inverse, pour une demande de connexion d'une requête dont la bande effective vaut α, on accepte celle-ci, si la quantité $W + \alpha$ est plus petite que la bande passante du nœud, sinon elle est rejetée.

Quand les trafics se distinguent par des niveaux de priorité (comme dans l'architecture de type Diffserv), les travaux sur le contrôle d'admission se ramènent essentiellement à supposer que la classe la plus prioritaire capte une portion fixe de la bande passante et à étudier ensuite le contrôle d'admission des autres trafics sur un nœud où la bande passante est réduite. Les travaux de Berger et Whitt illustrent ce type d'approche appelée habituellement *reduced service rate approximation* (RSR). Cette technique est connue pour être pertinente pour certaines disciplines de service comme WFQ *weighted fair queueing*. Dans le contexte envisagé ici, plusieurs études ont toutefois montré que ce type d'approximation pouvait conduire à sous-estimer la charge réelle du nœud et donc accepter trop de connexions qui n'auraient plus le niveau de qualité de service requis. Il est important de comprendre comment les niveaux de qualités de service peuvent être assurés et quand la propriété RSR est valide, ce qui conduit à une séparation virtuelle entre les différents trafics. À l'inverse, quand cette propriété n'est plus vérifiée, il s'agit de déterminer si le contrôle d'admission peut toujours s'effectuer de façon simple. Un travail préliminaire récent a montré que l'approximation RSR n'est pas valable sous certaines hypothèses de trafic et de priorité. Les travaux qui sont menés concernent à la fois les implications algorithmiques de ce type de résultat et l'étude des phénomènes responsables de l'échec de la RSR.

3.4. Allocation de bande passante

Key words: *Allocation de bande passante, Algorithmes MaxMin, Équité.*

Le thème de cette activité est l'allocation de bande passante dans un réseau transportant du trafic élastique contrôlé par TCP. Actuellement les flots se partagent la bande passante de façon égalitaire. L'implémentation de TCP (voir plus haut) est telle que, macroscopiquement, les ajustements se font sur les nœuds les plus chargés et, à ces nœuds la bande passante est équitablement répartie entre les messages. Si les mécanismes de ce type de politique ont l'avantage de réguler correctement, de façon distribuée le trafic, ils présentent l'inconvénient de ne pas utiliser pleinement la capacité du réseau. En effet, si par exemple une connexion traverse une série de $N - 1$ nœuds vides ayant bande passante maximum λ puis un n|oeud où passent M connexions, les mécanismes d'autorégulation feront que la connexion sera globalement transmise au taux λ/M à travers le réseau. Seulement une petite fraction de la capacité totale du réseau sera utilisée, λ/M par nœud au lieu de λ dans le cas idéal.

Le but de cette étude possibilité est d'augmenter l'utilisation de la capacité d'un réseau en modifiant les algorithmes de partage de bande passante. On se focalise sur la "physique" d'un réseau mettant en œuvre des politiques de partage de bande passante, ceci afin de dégager des heuristiques à l'aide de modèles mathématiques. Le cadre classique pour étudier le partage de bande passante dans les réseaux est celui des réseaux avec perte définis dans le livre de Kelly par exemple. Les études menées dans ce domaine ont surtout concerné des modèles où les messages sont transmis à des débits fixés à l'avance. Les résultats portent généralement sur l'évaluation des taux de perte ou de l'utilisation des liens du réseau (optimisation par des politiques de seuils ou *trunk reservation*). Les problèmes de reroutage des messages ont aussi fait l'objet d'analyses assez poussées tel le reroutage alternatif qui donne une meilleure occupation globale du réseau. Les questions de routage ne sont pas, pour l'instant, abordées.

Les études se font d'un point de vue macroscopique. Chaque connection TCP est vue de façon fluide et elle essaie d'écouler de façon continue une quantité x à travers le réseau. Cette approche ne considère donc pas la connection TCP au niveau microscopique, i.e. au niveau des transferts de paquets. Il s'agit de l'évaluation de plusieurs stratégies d'allocation de bande passante en particulier MaxMin. La politique MaxMin est vue comme une représentation macroscopique (i.e. fluide) de la façon dont TCP organise le trafic. Schématiquement l'allocation se fait sur le nœud le plus chargé et, sur celui-ci, la bande passante est

distribuée de façon équitable entre les connexions présentes, l'algorithme est ensuite répété en retirant les capacités allouées ainsi que les connexions concernées. Il faut noter que la topologie du réseau est un aspect très important de cette problématique. Cet algorithme est très difficile à évaluer qualitativement autrement que par des simulations. Nous nous intéressons à une variante, l'algorithme Min, dont les performances minorent celles de Maxmin. L'objectif actuel est d'essayer d'obtenir des résultats qualitatifs dans des configurations en surcharge de trafic.

4. Application Domains

4.1. Panorama

Les applications de nos travaux concernent la modélisation et l'étude des réseaux de télécommunication. Les principaux objectifs de RAP sont :

1. Le contrôle d'admission à l'entrée d'un réseau IP. Voir la section 3.3;
2. La métrologie. Voir la section 3.2;
3. L'allocation de bande passante à l'intérieur d'un réseau (réservation et équité). Voir la section 3.4.

5. Software

5.1. La plateforme ASIA

La plate-forme ASIA (*Accelerated Signalling for the Internet over ATM*) a été développée dans le cadre d'un projet RNRT (cf. http://www.telecom.gouv.fr/rnrt). Ce projet a été mené conjointement par France Telecom R&D (chef de file), Ericsson France, l'INRIA/IRISA et AIRTRIA qui est PME spécialisée dans le développement de logiciels pour les réseaux de télécommunications. Le projet a permis de mettre au point un réseau expérimental afin de démontrer la viabilité de certaines techniques pour écouler du trafic Internet sur ATM, tout en garantissant un certain niveau de qualité de service pour les applications. Les techniques utilisées dans ASIA sont :

- mise au point d'une plate-forme de médiation afin de permettre à un utilisateur de négocier de la qualité de service pour certains de ses flux, par exemple un flux vidéo;
- implantation de piles du protocole MPLS sur l'équipement AXD 312 de Ericsson;
- association dynamique de LSP (*label switched path*) créés par MPLS (en d'autre termes des connexions ATM sans débit) à des flux pour lesquels de la qualité de service a été négociée (essentiellement sous forme d'un débit minimum) et réservation de débit en temps réel sans latence pour l'application;
- renégociation du débit d'un LSP suivant un critère d'équité.

Le réseau expérimental ASIA a permis de tester les nouvelles architectures de réseaux dans le domaine des réseaux de nouvelle génération (NGN, *next generation network*). De plus, ASIA a validé le principe d'asservissement des connexions TCP par l'espacement de cellules ATM.

À l'avenir, ASIA devrait évoluer pour tenir compte des évolutions du réseau, en particulier de son architecture et des nouveaux services. Le réseau ASIA sert de banc de test à de nouvelles politiques d'acceptation de connexions (CAC) ainsi qu'à de nouveaux principes d'équité qui sont développés par l'équipe RAP . Par ailleurs, ASIA offre une plate-forme pour étudier l'intégration de flux temps réel et élastiques.

6. New Results

6.1. Analysis of Traffic Measurements

Participants: Nelson Antunes, Nadia Benazzouna, Christine Fricker, Fabrice Guillemin, Philippe Robert.

For traffic analysis, we adopt in this study a flow based approach and the popular mice and elephants dichotomy. A TCP flow is characterized by a sequence of packets characterized by four integers: source and destination addresses, source and destination ports. A mouse is a TCP flow with less than 20 packets, only the slow start phase of TCP protocol is used. On the contrary, due to their lengths, elephants share the remaining bandwidth becaus of the flow control mechanism of TCP. These two types of flows have therefore a completely different behavior from a modeling point of view.

Trace captures have been done by France Telecom R&D. TCP traffic has been collected on an Internet backbone link connecting different ADSL areas. A significant part of of it are p2p applications and hence large elephants.

6.1.1. Modeling ADSL traffic

An extensive statistical study of the traffic trace has been achieved. It has been observed the inter-arrival times of mice have an exponential distribution. Unfortunately, the mice arrival process is not Poisson as can be expected at first glance. This is due to the fact that mice are not actually independent but are sent by clumps.

But if mice of non p2p traffic with the same source address are aggregated, the stationary distribution of the number of macro-mice can then be described as Poisson. The distribution of the duration of these aggregated mice is Weibull. A complete description of the non p2p mice traffic (parameters of the distributions) has therefore been obtained.

For the p2p mice traffic, a second level of aggregation appears to be necessary. The reason is that the requests for a source at different destinations generate response messages. The conclusions are then identical as in the non p2p case.

For the elephant traffic, it has been observed (but never been mentioned in earlier studies) that the transmission of elephants is interrupted by time periods of several seconds. These parts of elephants are then considered as distinct elephants . Though the volume of a long flow has a Pareto distribution, it can be noticed that the distribution of the duration of these flows is Weibull.

A theoretical approximated model has been proposed. Indeed the mice processes are volumes in a $M/G/\infty$ where customers are characterized by their arrival time, their service durations and their profile. Moreover, the arrival rate is large so the processes can be approximated by a fluid approximation which leads to a Gaussian process. For the elephants, it is worth noticing that the $M/G/\infty$ queue is a good model due certainly to congestion on the access links. For the theoretical model, we have explicit expressions for the Laplace transform of the stationary rate and the stationary autocorrelation (even the transient ones). In the cases of specific distributions of the transmission duration of the flows, asymptotics for the autocorrelation have been obtained. This can be deduced from a general result by Borovkov and Iglehart. In the case of Poisson arrivals, an elementary proof in term of martingales of the heavy traffic limit in a $M/G/\infty$ queue has been derived.

This work [12] has been presented in an INRIA report and submitted to ICC.

6.1.2. Interaction of TCP flows

The integration of two types of flows sharing a channel is an important issue in the design of communication networks. It applies in the context of IP networks to the case of best effort traffics (TCP) and streaming traffics (UDP). It is also relevant to describe the interaction of short TCP flows and large TCP flows. (See the above section).

The basic model analyzed here consists in a bottleneck link with variable capacity receiving TCP connections. The varying capacity is due to the UDP traffic or the mice traffic depending on the model considered. The general idea being that, contrary to “greedy” traffic like UDP traffic or mice traffic, the large TCP flows adapt their throughputs to the state of the link.

The varying capacity is driven by a stationary Gaussian process, an Ornstein-Uhlenbeck process. This assumption is natural based on the observations of traffic measurements (See section above) and also, from a mathematical point of view, since it is known that the superposition of many small connections converges to such processes.

Up to now, the problem of expressing the invariant distribution of these systems is largely unsolved and known to be very difficult. Some earlier work by Nunez-Queija solves the problem in the context of a varying capacity driven by a Markov Modulated Point Process (MMPP). The result obtained are expressed in terms of complicated matrix expressions involving the (numerous) parameters of the MMPP. They are not easy to use in practice. This is also the reason why an Ornstein Uhlenbeck process has been taken, it is quite simple since it has only two parameters, the mean and the variance.

It has been chosen to study the case where the varying capacity oscillates around a fixed value μ. Perturbation methods have been used to describe the stationary behavior of such a system. To tackle this problem two approaches have been used. The first one which is analytic consists in expanding the solution of the PDE associated to the dynamic of the system with respect to the perturbation parameter. As a result a reduced load approximation has been proved when the capacity varies linearly with respect to the Gaussian process.

The other approach is probabilistic, it consists in analyzing the effect of the perturbation on one cycle of the Markov process. Under quite general assumptions, careful calculations lead to the expansion up to the second order of the stationary queue length. Our work is focusing now on a similar expansion but for the stationary sojourn time process.

6.2. Processor-Sharing disciplines with heavy tailed services

Participants: Fabrice Guillemin, Philippe Robert, Bert Zwart.

The processor-sharing paradigm has emerged as a powerful concept for modeling the flow-level performance of bandwidth-sharing protocols in communication networks. In this context, the driving random variables (especially service times) of PS models are often assumed to be heavy-tailed, reflecting the extreme variability of file transfers and session lengths. In view of this, several studies have focused on the difficult problem of analyzing the tail of the sojourn time distribution for the $M/G/1$ PS queue under heavy-tailed assumptions.

The framework considered here is a bottleneck link processing TCP connections as a processor-sharing queue with a possibly varying capacity, impatient flows and maximum number of connection at a given time. Our analysis consists in showing a general sufficient condition for RSR (Reduced Service Rate) approximation for these kind of queues. It is proved that, in this setting, a “big flow” receives a constant fraction of the total capacity of the link capacity. These results apply to class of state- dependent PS queues, PS queues with finite buffers and/or impatience. Based on these results, an admission control policy have designed in an earlier work with jacqueline Boyer.

6.3. Multi-class queues

Participants: Abdel Ben Tahar, Philippe Robert.

The analysis of a transient multi-class queue has been achieved. Customers enter the queue in the FIFO order and achieve J feed-backs before leaving definitely the queue. The distribution of the service received at the jth feedback depends on j. When the load is greater than one, limit results have been proved for the total number of customers in each class. The techniques involve the use of branching processes. A study of the service in random order discipline has been also done in the same context.

6.4. Book

Participant: Philippe Robert.

Philippe Robert has translated his book “Réseaux et files d’attente: méthodes probabilistes”, it appeared as “Stochastic networks and queues” in the Stochastic Modelling and Applied Probability Series by Springer Verlag New-York.

7. Contracts and Grants with Industry

7.1. Contrats industriels (Nationaux, Européens)

Participation to the CRE with France Telecom “Bandwidth Allocation in the Internet”. This contract is for two years.

Participation to the RNRT project on the measurements in the Internet. Four years contract ending in 2004.

Participation to the ACI GAP “Graphs, Algorithms and Probability” on the algorithms of peer to peer networks. Participants: INRIA, LIAFA and LRI. Three years contract.

Participation to the ACI Xtrems on the rare events in the modeling of the Internet. Participants: ENST, INRIA and University of Evry. Three years contract.

Participation to E-Next, a network of excellence of EC.

8. Other Grants and Activities

8.1. Actions nationales

Philippe Robert et *Fabrice Guillemin* are participating to the “Action Spécifique Métrologie”. The other members are Pascal Abry (ENS-Lyon), Daniel Kofman (ENST), Philippe Owezarski (LAAS) and Kavé Salamatian (Paris VI).

8.2. Actions financées par la Commission Européenne

RAP is participating to the E-next network of excellence of EC. This network involves many reasearch teams throughout Europe. In France, participants include LIP6, INRIA-Sophia, LAAS,...This network is a continuation of the efforts of RAP team in the domain of traffic measurement.

8.3. Accueils de chercheurs étrangers

RAP team has received the following people:
Predrag Jelenkovic (Columbia University), Kavita Ramanan (Carnegie Mellon University), Ahmed Kharroubi (University of Casablanca) and Bert Zwart (Eurandom).

9. Dissemination

9.1. Animation de la communauté scientifique

Fabrice Guillemin participated to the program comitee of INFOCOMM’2004 and ICC’2004 and to the CR2 jury at INRIA-Sophia.

Philippe Robert has been the referee for the PhD theses by A. Proutière from École Polytechnique and L. Rabehasaina from the University of Rennes. He participated to the CR2 jury at INRIA-Rocquencourt.

Philippe Robert is “Professeur Chargé de Cours à l’École Polytechnique” in the department of applied mathematics. He is in charge of the lectures on the mathematical modeling of networks.

9.2. Enseignement universitaire

Christine Fricker gives DEA lectures “Stochastic Processes” at the University of Versailles St-Quentin.

Philippe Robert gives DEA lectures “Stochastic Networks” in the laboratory of the Probability of the University of Paris VI. He is also giving lectures in the “Majeure de Mathématiques Appliquées et d’Informatique” on Networks and Algorithms at the École Polytechnique.

9.3. Participation à des colloques, séminaires, invitations

Christine Fricker and *Fabrice Guillemin* were at the Eighteenth International Teletraffic Congress (ITC-18) from August 31st to September 5th, 2003 held in Berlin, Germany.

Christine Fricker and *Philippe Robert* were at the ARC TCP meeting from November 5th to November 7th at ENS Paris .

Christine Fricker was from April 9th to April 11th and from July 15th to July 18th, at France Telecom R&D in Lannion, France.

Fabrice Guillemin has been invited at the RHDM conference on the quality of service in the Internet.

Philippe Robert has been visiting MISTRAL team in Sophia from 02/04 to 02/05. He was at the Infocom conference in San Francisco from 03/31 to 04/03.

He was invited by

- K. Sigman at the Applied probability day conference at the CAP in Columbia University (New York) from 04/04 to 04/08,
- S. Asmussen at the University of Aarhus from 04/21 to 04/25,
- L. Goldberg at the University of Durham for the Conference on Markov Chains from 07/25 to 07/31.
- O. Boxma at Eurandom (Eindhoven) at the workshop on heavy traffic.

He visited France Telecom Lannion from 07/15 to 07/16. He was invited speaker at the conference “Un siècle de Mathématiques à Nancy” at Nancy from 10/23 to 10/24

10. Bibliography

Books and Monographs

[1] P. ROBERT. *Stochastic Networks and Queues.* series Stochastic Modelling and Applied Probability Series, volume 52, Springer, New-York, June, 2003.

Articles in referred journals and book chapters

[2] C. FRICKER, P. ROBERT, D. TIBI. *A degenerate central limit theorem for single resource loss systems.* in « Annals of Applied Probability », number 2, volume 13, 2003, pages 561–575.

[3] F. GUILLEMIN, R. MAZUMDAR, A. DUPUIS, J. BOYER. *Analysis of the fluid weighted fair queueing system.* in « Journal of Applied Probability », volume 40, 2003, pages 180–199.

[4] F. GUILLEMIN, R. MAZUMDAR. *Rate conservation laws for multidimensional processes of bounded variation with applications to priority queueing systems.* in « Methodology and Computing in Applied Probability », 2004, To appear.

[5] F. GUILLEMIN, D. PINCHON. *Analysis of the weighted fair queueing system with two classes of customers with exponential service times..* in « Journal of Applied Probability », 2004, To appear.

[6] F. GUILLEMIN, P. ROBERT, B. ZWART. *Tail Asymptotics for Processor Sharing Queues.* in « Journal of Applied Probability », 2003, submitted.

[7] I. MITRANI, P. ROBERT. *On the ASTA Property in a Feedback Processor-Sharing Queue.* in « Performance Evaluation », 2003, Submitted.

Publications in Conferences and Workshops

[8] N. B. AZZOUNA, F. GUILLEMIN. *Analysis of ADSL traffic on an IP backbone link.* in « Proc. Globecom 2003 », December, 2003.

[9] J. BOYER, F. GUILLEMIN, P. ROBERT, B. ZWART. *Heavy tailed M/G/1-PS queues with impatience and admission control in packet networks.* in « Infocomm'2003 », San Francisco, USA, 2003.

[10] V. DUMAS, F. GUILLEMIN, P. ROBERT. *Admission Control of leaky bucket regulated sources in a queueing system with priority.* in « ITC'18 », Berlin, September, 2003.

[11] F. GUILLEMIN, N. LIKHANOV, R. MAZUMDAR, C. ROSENBERG. *Buffer overflow bounds for multiplexed regulated traffic streams.* in « Proc. ITC'18 », September, 2003.

Internal Reports

[12] N. BEN AZZOUNA, C. FRICKER, F. GUILLEMIN. *Analysis of ADSL traffic on an IP backbone link.* Technical report, number 4909, INRIA, august, 2003, http://www.inria.fr/rrrt/rr-4909.html.

Bibliography in notes

[13] T. BONALD, S. OUESLATY-BOULAHIA, J. ROBERTS. *QOS is still an issue: we need a new paradigm.* 2002, France Telecom technical report.

[14] V. DUMAS, F. GUILLEMIN, P. ROBERT. *Limit results for Markovian models of TCP.* in « Globecom'01, IEEE Global Telecommunications Conference », San Antonio, Texas, November, 2001.

[15] A. GAĬRAT, V. MALYSHEV, M. V. MEN'SHIKOV, K. PELIKH. *Classification of Markov chains describing the evolution of random strings.* in « Russian Mathematical urveys », number 2, volume 50, 1995, pages 237–255.

[16] L. MASSOULIÉ, J. ROBERTS. *Bandwidth sharing: Objectives and algorithms.* in « INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies », pages 1395–1403, 1999.