INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Équipe Rap

Réseaux, Algorithmes et Probabilités

Rocquencourt

THÈME 1B

Table des matières

1. Composition de l'équipe **1**
2. Présentation et objectifs généraux **1**
3. Fondements scientifiques **1**
3.1. Méthodes de renormalisation 1
3.1.1. Les insuffisances du cadre actuel 3
3.2. Métrologie 4
3.3. Contrôle d'admission 4
3.3.1. Contrôle d'admission du trafic élastique 5
3.3.2. Contrôle d'admission du trafic prioritaire 5
3.4. Allocation de bande passante 6
4. Domaines d'application **7**
4.1. Panorama 7
5. Logiciels **7**
5.1. La plateforme ASIA 7
6. Résultats nouveaux **8**
6.1. Contrôle d'admission du trafic élastique 8
6.2. Interaction des flots TCP sur des liens ADSL 8
6.2.1. Modélisation du trafic des souris 8
6.2.2. Intégration des souris et des éléphants 9
6.3. Étude de TCP au niveau paquet 9
7. Contrats industriels **10**
7.1. Contrats industriels (Nationaux, Européens) 10
8. Actions régionales, nationales et internationales **10**
8.1. Actions nationales 10
8.2. Actions financées par la Commission Européenne 10
8.3. Accueils de chercheurs étrangers 11
9. Diffusion des résultats **11**
9.1. Animation de la communauté scientifique 11
9.2. Enseignement universitaire 11
9.3. Participation à des colloques, séminaires, invitations 11
10. Bibliographie **12**

1. Composition de l'équipe

Responsable scientifique
Philippe Robert [DR]

Responsable permanent
Fabrice Guillemin [France Télécom R&D, Lannion]

Assistante de projet
Virginie Collette [TR]

Personnel Inria
Christine Fricker [CR]

Collaborateurs extérieurs
Alain Dupuis [France Télécom R&D, Lannion]
Jacqueline Boyer [France Télécom R&D, Lannion]
Danielle Tibi [Université Paris VII]

Chercheur post-doctorant
Albertus Zwart [Jusqu'au 1er Août]

Doctorants
Mostapha Haddani [Jusqu'au 30 juin]
Nadia Benazzouna [France Telecom R&D Lannion]

2. Présentation et objectifs généraux

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 [11]). 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. Voir les articles [5] et[7] qui illustrent ce type d'approche.

3. Fondements scientifiques

3.1. Méthodes de renormalisation

Mots clés : *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 $\bar{X}$ associé est donné par

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

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 $(\bar{X}(x,t))$ 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 $\bar{X}$ 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 $t \longrightarrow 1 + \delta t$. 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 $t \longrightarrow 1 + \delta t$. 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 qui est généralement méconnu et qui 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. Dans Dantzer *et al.* [14] nous montrons que, dans un cadre simple, un algorithme d'allocation de bande passante naturel a des limites fluides oscillantes dirigées par une chaîne de Markov à espace d'états fini. Récemment, dans un cadre très différent, Fricker *et al.* [6], nous montrons que les limites fluides des très classiques réseaux avec perte gardent en fait une composante aléatoire, contredisant ainsi une ancienne conjecture dans ce domaine. Ces questions constituent un domaine d'étude très important du comportement des réseaux.
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.* [16] 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 [4] introduisent plusieurs notions qui nous 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

Mots clés : *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 [15] 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

Mots clés : *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 [17]. 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

Mots clés : *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. Domaines d'application

4.1. Panorama

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

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. Logiciels

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. Résultats nouveaux

6.1. Contrôle d'admission du trafic élastique

Participants : Jacqueline Boyer, Christine Fricker, Fabrice Guillemin, Philippe Robert, Bert Zwart.

Le contexte est celui de la partie 3.3.1, l'étude [11] qui a été menée à ce jour a consisté à représenter le trafic des éléphants par un flot de Poisson dont les services ont des distributions de Pareto, i.e. dont la queue de distribution décroît de façon polynomiale. Ce flot décrit les arrivées des éléphants sur le lien. De plus, la durée d'impatience d'une connexion est supposée être proportionnelle à la taille transférée, i.e. $I = Cx$ où x est la taille, I l'impatience, et C est une constante plus grande que 1. Plus le fichier est important et plus la connexion est « <patiente ». Si $L(t)$ est le nombre de connexions actives à l'instant t, la quantité

$$S(x) = \int_0^x \frac{1}{1 + L(s)}\, ds$$

représente le temps nécessaire pour transférer un fichier de taille x. La probabilité qu'il n'y ait pas interruption est la probabilité que $S(x) < I = Cx$. Nous avons proposé un contrôle d'admission, par le biais d'un buffer virtuel de taille N, qui consiste à rejeter systématiquement tous les éléphants dès que ce buffer est rempli. Le principal résultat montre qu'il y a deux seuils N_0 et N_1 pour la taille du buffer virtuel avec les propriétés respectives suivantes :

- Si $N \leq N_0$, la discipline est conservative. Les pertes dues au rejet ont complètement remplacé les pertes dues à l'impatience, les simulations montrent de plus que, si $N = N_0$, la disparition de l'*overhead* due à l'impatience a entraîné un gain de capacité pour le système ;
- Si $N \leq N_1$, alors un éléphant s'il est admis, sa probabilité d'impatience tend vers 0 quand sa taille tend vers l'infini. Sous ces conditions, l'algorithme traite de façon équitable les très gros transferts ;
- À l'inverse, si $N > N_1$, un très gros éléphant est impatient avec une probabilité proche de 1.

Dès que le seuil est en-dessous de N_1 les « gros » transferts ne sont plus pénalisés. Ce résultat peut être interprété comme une propriété d'équité de l'algorithme de contrôle d'admission.

Par la même occasion, ce travail a permis de montrer qu'un résultat important de charge réduite équivalente était vrai dans un cadre général. Si S est le temps de séjour d'un client dans une file d'attente processor-sharing avec impatience, la queue de distribution de S est équivalente à la queue de distribution du service modulo un coefficient multiplicatif.

6.2. Interaction des flots TCP sur des liens ADSL

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

Le but de ce travail est d'analyser le trafic du réseau Internet dans le but d'avoir des modèles réalistes de trafic. D'après Floyd and Paxson, le trafic Internet est vu, en première approximation, comme composé de souris et d'éléphants. Les deux types de flots ont des comportements très différents. Le trafic des souris est un trafic qui ne subit pas le contrôle de congestion, mais qui occupe une partie de la bande passante. Les éléphants se partagent ce qui reste, tout en étant régulés par le protocole TCP. Il convient d'isoler ces deux types de trafic pour les caractériser. Le captage des traces sur le réseau a été fait par France Telecom R&D sur du trafic ADSL.

6.2.1. Modélisation du trafic des souris

Un flot est une suite de paquets caractérisée par la donnée de quatre entiers : une adresse et un numéro de port pour la source et pour la destination. Une souris est définie comme un flot TCP de moins de n paquets où n est choisi pour que le flot ne sorte pas de la phase de slow-start ($n = 10$ ou 20). Les observations portent sur la caractérisation du trafic des souris. Au vu des mesures, les inter-arrivées des souris ont une loi qui est

remarquablement approximée par une loi exponentielle. Ce qui ne veut pas dire que le trafic des souris est de Poisson ! La durée D des souris semble par contre sous-exponentielle (loi de Weibull), i.e.

$$\mathbb{P}(D \geq x) \sim Ce^{-\alpha x^{\beta}},$$

où $\beta \sim 0.85$ et $\alpha \sim 0.55$. Le nombre de paquets d'un flot court dépend de la durée du flot. Cette dépendance capte le fait que les temps de transmission aller et retour (RTT) sont différentes d'un flot à l'autre. En fait, ce nombre est très concentré sur la valeur de 2 paquets et l'idée est, de prendre pour une durée s, le débit c/s où c est une constante (ou une loi que l'on choisira).

Pour résumer, cela conduit à un modèle où les souris (clients) arrivent de façon poissonnienne et génèrent des arrivées de paquets (modèle discret) ou en première approximation un débit moyen (modèle fluide) pendant une durée de transmission (ou service). On peut même envisager des « profils » fluides pour les flots plus compliqués que des rectangles (*shot noise*) pour tenir compte de l'algorithme de *slow-start* de TCP. Dans le cadre des processus de Poisson sur $\mathbb{R}^{+\,3}$, le modèle fluide avec profil s'avère facile à analyser. C'est une généralisation de la file $M/G/\infty$. On obtient des formules explicites de la transformée de Laplace du débit transitoire et stationnaire, la fonction de covariance transitoire et stationnaire. Les asymptotiques de l'autocorrélation, dans le cas de différentes lois de longueurs de souris, peuvent être alors obtenues : la décroissance de l'autocorrélation du débit suit celle de la longueur (en particulier Weibull). Quand le taux d'arrivée augmente (*heavy traffic*), le débit converge vers un processus gaussien limite dont l'autocorrélation est celle du processus initial.

On compare actuellement ces résultats théoriques avec les résultats de mesures discrètes, fluides et de simulations. Le problème est de se débarrasser des parties bruits induites par la discrétisation du temps et les arrivées discrètes des paquets. On se propose si nécessaire de faire l'étude d'un modèle discret de flots.

Ce travail a aussi permis d'avoir une preuve élémentaire en terme de martingales du *heavy traffic* d'une file $M/G/\infty$, découlant d'un résultat plus général dû à Borovkov et Iglehart.

6.2.2. Intégration des souris et des éléphants

Il s'agit d'étudier l'influence du trafic des souris sur les performances du trafic des éléphants. Les deux types de clients coexistent dans une file à capacité limitée. Les souris y sont servies en parallèle et sont perdues si la capacité est dépassée, les éléphants sont servis avec la capacité restante par un serveur *processor sharing* à taux de service variable. On renormalise le processus du nombre des souris et des éléphants dans le cas où la capacité de la file et les taux d'arrivées sont multipliés par N. Les limites du processus renormalisé sont obtenues au premier ordre (limite fluide) et au deuxième ordre (diffusion). On étudie aussi le cas de trafic chargé, quand le débit des souris est inférieur à C et le débit global tend vers C.

6.3. Étude de TCP au niveau paquet

Participants : Fabrice Guillemin, Philippe Robert, Bert Zwart.

Cette partie concerne l'étude de la transmission des paquets par une connexion TCP acheminant un fichier de taille infinie à travers le réseau. Plus précisément, nous avons étudié le comportement asymptotique, quand le taux de perte devient très petit, de la taille de la fenêtre de congestion associée à la connexion. En utilisant un modèle très simple où le réseau perd des paquets de façon aléatoire, nous avons montré dans [5] que la suite des carrés des tailles des fenêtres de congestion convenablement renormalisées formait une suite auto-régressive. Ce résultat est, à notre connaissance, le premier qui mette en évidence cette remarquable propriété.

Les expressions explicites du débit de la connexion TCP et la densité de la distribution à l'équilibre de la taille de la fenêtre de congestion ont été ainsi obtenues. La queue de distribution de celle-ci décroît en $\exp(-ax^2)$ et non exponentiellement comme pouvaient le suggérer certains modèles utilisés précédemment dans la littérature. Qualitativement, cela implique que le nombre de grandes fenêtres de congestion est surestimé par ces modèles. Il faut noter que nos résultats intègrent le fait que, en réalité, le protocole a une fenêtre de congestion de taille maximale, ce qui n'est pas pris en compte par la plupart des modèles actuels.

Les expressions analytiques des résultats sont sensiblement plus compliquées par rapport au cas de la taille infinie de la fenêtre de congestion maximale.

Cette étude a été poursuivie et amplifiée cette année [7]. En utilisant l'analyse de mesures menée par Paxson et Zhang sur les processus de pertes de paquets sur Internet, nous avons été en mesure de proposer un modèle réaliste des pertes de paquets. L'ingrédient essentiel, qui n'était pas dans notre étude précédente, ni d'ailleurs dans toutes les études analytiques antérieures, est que les pertes de paquets sont corrélées. Si un paquet est perdu, c'est en général dû à un buffer de routeur saturé, et donc, d'autres pertes consécutives sont prévisibles.

En considérant des pertes corrélées, toujours en renormalisant la taille des fenêtres de congestion avec le taux de perte, nous avons montré que les théorèmes limites obtenus dans [5] s'étendaient à ce cas. L'étude des distributions explicites s'est avérée beaucoup plus délicate. Si la transformée de Laplace de la taille limite de la fenêtre de congestion est explicite, son utilisation pratique est plus délicate car, à la différence de [5], elle ne s'inverse pas toujours. Sur le plan analytique, le cadre naturel de l'étude est celui des q-fonctions hypergéométriques. Sur le plan probabiliste, le cadre est celui des intégrales exponentielles de processus de Lévy. Dans notre cas, les processus de Lévy sont des processus de Poisson composés. Quand le processus de Lévy est un brownien avec dérive, cela rejoint les études menées ces cinq dernières années par Yor et ses collaborateurs en mathématiques financières. Il est intéressant de constater cette proximité inattendue de modèles mathématiques d'objets a priori différents.

Le travail a aussi consisté à quantifier l'impact de la corrélation sur le débit de la connexion TCP. Pour ce faire, nous avons obtenu une formule explicite du débit via un résultat original sur les moments fractionnaires des intégrales exponentielles. La principale conclusion est que, à taux de perte constant, plus le processus de perte est variable, meilleur est le débit. En particulier, le modèle étudié en détail dans [5] est le cas le pire, donc si le modèle des pertes de paquets indépendantes est discutable, son utilisation revient cependant à donner une borne inférieure sur le modèle réel.

7. Contrats industriels

7.1. Contrats industriels (Nationaux, Européens)

Ch. Fricker et Ph. Robert participent à la consultation thématique de France Telecom R&D sur l'optimisation de la gestion du trafic TCP dans un réseau. (Voir la partie résultats nouveaux pour la description de cette action). Ce contrat est d'une durée de deux ans et se termine en décembre 2002.

Une nouvelle proposition « Allocation de bande passante sur Internet » a été soumise, si celle-ci est retenue, ce travail se fera également en collaboration avec l'équipe de J. Roberts à France Telecom R&D.

C. Fricker et Ph. Robert participent au projet RNRT Métropolis sur l'utilisation de la métrologie dans l'étude des réseaux IP (Voir la description dans la section résultats nouveaux). Ce contrat est d'une durée de trois ans.

8. Actions régionales, nationales et internationales

8.1. Actions nationales

Philippe Robert et *Fabrice Guillemin* participent au comité de pilotage de l'Action Spécifique Métrologie. Les autres membres sont Pascal Abry (ENS-Lyon), Daniel Kofman (ENST), Philippe Owezarski (LAAS) et Kavé Salamatian (Paris VI).

8.2. Actions financées par la Commission Européenne

L'avant-projet RAPest avec le projet ALGO pour une période de trois ans, 2000-2003, l'une des composantes du projet Esprit BRA Alcom-FT (*Algorithms and Complexity-Future Technologies*) de l'Union Européenne avec neufs partenaires : University of Aarhus, Polytechnic University of Catalunya, University of Cologne,

Max-Planck-Institut für Informatik, University of Paderborn, Computer Technology Institute (Patras), University of Roma « La Sapienza », University of Utrecht, University of Warwick. L'objectif affiché est la découverte de nouveaux concepts algorithmiques et l'identification des algorithmes clefs transverses à de nombreuses applications. Quatre directions de travail ont été identifiées : (i) ensembles de données massifs ; (ii) systèmes de communication complexes ; (iii) optimisation en production et planification ; (iv) recherches méthodologiques et expérimentales en algorithmique. Les travaux du projet se situent principalement dans les axes (ii) et (iv).

8.3. Accueils de chercheurs étrangers

L'avant-projet RAP a reçu les visites de *Christian Gromoll* (Eurandom) du 4 au 8 mars, *Nelly Litvak* (Eurandom) du 21 au 22 mars, *Michel Mandjes* (CWI) du 8 au 12 avril, *Karl Sigman* (Columbia University) le 18 avril, *Kavita Ramanan* (Lucent) du 27 août au 2 septembre, *Isi Mitrani* (Université de Newcastle) du 2 au 5 septembre.

9. Diffusion des résultats

9.1. Animation de la communauté scientifique

Fabrice Guillemin a été membre du comité de programme d'INFOCOMM'2003, ECUNM02 et de l'issue spéciale de TSI sur les réseaux.
Philippe Robert a été membre du comité de programme de la conférence « Mathematics and Computer Science », à l'Université de Versailles St-Quentin, du 18 au 20 septembre. *Philippe Robert* a été élu membre du groupe WG 7.3 *Computer Performance Modeling and Analysis*, de l'IFIP. *Philippe Robert* a été le rapporteur des thèses de S. El Merzouki, de l'Université de Rouen et de G. Regnié, de l'Université de Paris VI.

9.2. Enseignement universitaire

Philippe Robert donne un cours de DEA intitulé « Processus stochastiques » au DEA Maths-Info de l'Université de Versailles St-Quentin, un cours de DEA intitulé « Réseaux et protocoles de télécommunication : modèles probabilistes » au DEA de probabilités du laboratoire de probabilités de l'Université Paris VI, ainsi qu'un cours d'option à l'ENST, sur les modèles probabilistes du protocole TCP. Il donne un cours de modélisation de réseaux, en maîtrise d'informatique, à l'Université de Cergy-Pontoise.

9.3. Participation à des colloques, séminaires, invitations

Un groupe de travail sur le problème de Skorohod a été organisé aux mois de mai et juin. L'objectif de ce groupe de travail était de faire le point en quatre exposés par Christine Fricker, Danielle Tibi, Philippe Robert et Bert Zwart sur ces questions.
Christine Fricker a participé aux journées ARC TCP du 16 au 17 mai à Sophia-Antipolis et à la conférence ITC « Internet traffic engineering and traffic management » du 22 au 24 juillet 2002 à Wurzburg. Elle a passé une semaine à France Telecom R&D, Lannion, du 8 au 12 juillet.
Fabrice Guillemin a participé à la conférence INFOCOMM'2002 à New-York, du 23 au 27 juin, il a également participé au workshop ITC *Specialist Seminar* qui s'est tenu à Wurzburg en Allemagne, du 21 au 24 juillet.
Philippe Robert a donné un exposé sur les algorithmes de contrôle de la congestion dans les réseaux à l'INPG le 17 janvier. Il a rendu visite à A. El Kharroubi du département de mathématiques de l'Université Ain Chok de Casablanca au Maroc, du 22 janvier au 27 janvier. Deux conférences y ont été données une sur les modèles markoviens de TCP et l'autre sur les problèmes d'allocation de bande passante. *Philippe Robert* a donné un séminaire « Modèles markoviens de TCP » au projet Algo le 11 février, à la journée ARC-TCP à l 'INRIA-Sophia, le 16 mai, au département de mathématiques de l'Université de Dijon, le 21 mai et au département de mathématiques de l'Université de Nancy, le 13 juin. Il a également participé à la revue du projet européen ALCOM-FT qui s'est tenu à Warwick, GB, du 4 au 7 juillet. *Philippe Robert* a rendu visite à A. Rybko et

N. Vvedenskaya à l'IPPI à Moscou, Russie, du 22 au 26 juillet. *Philippe Robert* a été invité au workshop *Analysis and optimisation of stochastic networks with application to telecommunications and manufacturing* à Eindhoven, Pays-Bas, du 7 au 9 novembre.
Bert Zwart a participé au workshop DYNSTOCH sur les dépendances à long terme, queues de distributions lourdes et les événements rares à Copenhague, Danemark, du 6 au 18 mai. Il a été invité par Thomas Mikosch au département de mathématiques actuarielles la semaine suivante. Il a participé au workshop *Modern problems in applied probability* à Edinburgh, Royaume-Uni, du 21 au 29 août.

10. Bibliographie

Articles et chapitres de livre

[1] O. BOXMA, Q. DENG, B. ZWART. *Waiting-time asymptotics for the $M/G/2$ queue with heterogeneous servers.* in « Queueing Systems », volume 40, 2002, pages 5-31.

[2] J. BOYER, F. GUILLEMIN. *Spectral analysis of the M/M/1 queue with processor sharing.* in « Queueing Systems », numéro 4, volume 39, 2001, pages 377-397.

[3] J.-F. DANTZER, V. DUMAS. *Stability analysis of the Cambridge ring.* in « Queueing Systems. Theory and Applications », numéro 2, volume 40, 2002, pages 125-142.

[4] J.-F. DANTZER, P. ROBERT. *Fluid limits of string valued Markov processes.* in « Annals of Applied Probability », numéro 3, volume 12, 2002, pages 860-889.

[5] V. DUMAS, F. GUILLEMIN, P. ROBERT. *A Markovian analysis of Additive-Increase Multiplicative-Decrease (AIMD) algorithms.* in « Advances in Applied Probability », numéro 1, volume 34, 2002, pages 85-111.

[6] C. FRICKER, P. ROBERT, D. TIBI. *A degenerate central limit theorem for single resource loss systems.* To appear in the "Annals of Applied Probability".

[7] F. GUILLEMIN, P. ROBERT, B. ZWART. *AIMD algorithms and exponential functionals.* 2002, To appear in the "Annals of Applied Probability".

[8] W. SCHEINHARDT, B. ZWART. *A tandem fluid queue with gradual input..* in « Probability in the engineering and informational sciences », volume 16, 2002, pages 29-45.

Communications à des congrès, colloques, etc.

[9] F. GUILLEMIN, N. LIKHANOV, R. MAZUMDAR, C. ROSENBERG. *Extremal traffic and bounds on the mean delay of multiplexed regulated traffic streams.* in « INFOCOM'2002 », juin, 2002.

[10] F. GUILLEMIN, P. ROBERT, B. ZWART. *Performance of TCP in the presence of correlated packet loss.* in « 15th ITC Specialist Seminar on Internet Traffic Engineering and Traffic Management », Wurzburg, juillet, 2002.

Rapports de recherche et publications internes

[11] J. BOYER, F. GUILLEMIN, P. ROBERT, B. ZWART. *Heavy tailed M/G/1-PS queues with impatience and admission control in packet networks.* rapport technique, numéro 4536, INRIA, septembre, 2002, http://www.inria.fr/rrrt/rr-4536.html.

[12] P. JELENKOVIC, P. MOMCILOVIC, A. ZWART. *Reduced Load Equivalence under Subexponentiality.* rapport technique, numéro 4444, INRIA, 2002, http://www.inria.fr/rrrt/rr-4444.html.

Bibliographie générale

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

[14] J.-F. DANTZER, M. HADDANI, P. ROBERT. *On the stability of a bandwidth packing algorithm.* in « Probability in the Engineering and Informational Sciences », numéro 1, volume 14, 2000, pages 57-79.

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

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

[17] 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.