Partage de Charge et Ingénierie de Trafic dans les Réseaux MPLS

Ramon Casellas

Thèse

présentée pour obtenir le grade de docteur
de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Informatique et Réseaux

M. Ramon CASELLAS

Partage de Charge et Ingénierie de Trafic dans les Réseaux MPLS

soutenue le 25 novembre 2002 devant le jury composé de

Annie Gravey	**Président**
Ravi R. Mazumdar	**Rapporteurs**
Philippe Nain	
James Roberts	**Examinateurs**
Samir Tohmé	
Jean-Louis Rougier	
Daniel Kofman	**Directeur de thèse**

A la mémoire de mon père, Ramon Casellas Fornés,
A ma mère, Araceli Regi i de Luque,
A mon frère, Toni Casellas Regi,
Merci.

Remerciements

Cette thèse n'aurait jamais vu le jour sans l'aide et le soutien d'un certain nombre de personnes auxquelles j'aimerais exprimer ici toute ma reconnaissance. Je voudrais tout d'abord remercier mon directeur de thèse, M. Daniel Kofman, pour sa confiance renouvelée et ses encouragements continus tout au long de ces trois années. Ses commentaires et ses remarques m'ont été très précieux et m'ont aidé à développer un esprit critique indispensable pour mener à bien un tel travail.

Je suis profondément redevable aux membres de mon jury : merci à Mme. Annie Gravey, directrice du département Informatique de l'ENST Bretagne, dont les commentaires toujours pertinents et la jovialité ont souvent relancé ma motivation ; merci à M. Ravi R. Mazumdar, Professeur à l'Université de Purdue, dont les commentaires et les remarques m'ont été très utiles pour le développement de la dernière partie de ce document, et qui a accepté d'être rapporteur de mon travail, et à M. Philippe Nain, directeur de recherche à l'INRIA, qui a également bien voulu être rapporteur du document.

M. Jim Roberts, de France Télécom Recherche et Développement m'a fait l'honneur de bien vouloir faire partie de ce jury ; M. Jean-Louis Rougier a été un excellent collègue et un très bon ami. Je le remercie d'avoir également accepté de faire partie de ce jury. Finalement, merci à M. Samir Tohmé, responsable du groupe Réseau Haut Débit du département INFRES de l'ENST, pour avoir accepté de faire partie de mon jury de thèse.

Dans le contexte de mes travaux de recherche, j'ai eu des discussions très enrichissantes avec plusieurs membres du département INFRES, notamment MM. Laurent Decreusefond et Hayri Korezlioglu ; qu'ils en soient grandement remerciés.

MM. François Baccelli, Laurent Massoulié et Jean Mairesse, responsables du DEA Probabilités Appliquées et Processus Ponctuels à l'Université Paris VI, m'ont facilité le premier contact avec la théorie des grandes déviations et son application aux réseaux de files d'attente. Je leur en suis sincèrement reconnaissant.

Merci à M. Anthony Busson, colocataire de bureau et ami, qui a été parfois obligé de me rappeler qu'une probabilité ne peut être supérieure à un, et à M. Christian Roche qui s'est gracieusement chargé de la lourde tâche de relire les versions préliminaires de ce document et de m'aider à aplanir mes difficultés avec la langue de Molière. Il se peut qu'un certain nombre de fautes de frappe soient encore présentes dans la version finale de ce document. Je fais appel à ce sujet à l'indulgence du lecteur.

Lors de mon séjour à l'ENST j'ai eu la chance de côtoyer un grand nombre d'enseignants-chercheurs et de thésards d'horizons divers avec lesquels j'ai partagé de très bons moments : M. Sergio Beker, M. Philippe Monnier, M. Philippe Martins, Mlle. Nadia Boukhatem, M. Gwendal Legrand et tant d'autres, trop nombreux pour être tous cités ici.

Finalement, je tiens à remercier mes amis à Barcelone, Juan et Ramon San-Martin et Sergio Cormano, qui m'ont soutenu et encouragé pendant tout mon séjour en France. Merci à Jesús García et

Célia Costéja, mes chers colocataires, qui ont supporté mes moments de fatigue et de pessimisme. Enfin, un grand merci à ma mère et à mon frère pour leur affection qui m'a été extrêmement précieuse durant ces années loin d'eux.

Résumé

Dans cette thèse, nous nous sommes intéressés à l'optimisation du partage de charge dans un réseau supportant le routage à la source. Une modélisation générique en files d'attente alimentées par un trafic caractérisé par sa bande passante effective et l'utilisation de la théorie des Grandes Déviations nous a permis de déduire de règles d'ingénierie dans divers contextes en optimisant des fonctions de coût qui reflètent les besoins des réseaux opérationnels. Des propriétés structurelles sur les politiques optimales ont été démontrées pour des cas particuliers mais importants dans le domaine de l'ingénierie du trafic.

La variabilité de la capacité des chemins d'un réseau a été intégrée à l'aide du concept de capacité effective. Nous avons mis en évidence qu'un dimensionnement basé sur une capacité moyenne peut s'avérer sous optimal et nous avons quantifié cela. Ainsi, une approche d'ingénierie de trafic adaptative a été proposée en faisant évoluer le partage en fonction des mesures réalisées sur le réseau.

Dans la dernière partie de ce travail nous avons regardé le réseau dans sa globalité, modélisant les interactions entre l'ensemble des couples entrée-sortie, nous permettant de mieux optimiser le réseau mais au coût d'une grande complexité de calcul.

Abstract

In this dissertation, we propose and optimize load sharing mechanisms adapted to a network supporting source routing. A generic model using queues fed by traffic which is represented by its effective bandwidth is defined, and the use of the large deviations theory allows us to obtain traffic engineering rules and guidelines for several scenarios by optimizing cost functions that reflect the needs of operational networks. Structural properties about optimal load sharing policies have been obtained in particular but important cases with regard to traffic engineering.

The time varying property of the end to end capacity of network paths has been taken into account by means of what we call effective capacity. We have shown that a load sharing based on average values may not be optimal, and we have quantified this. In this sense, we have also proposed and adaptive measurement based load sharing mechanism.

In the last part of our work, we have extended our model in order to consider the whole network, taking into account the interactions between all origin-destination couples, allowing a better optimization of the network but with a high computation complexity.

Administrativia

Le présent document est le manuscrit de thèse pour obtenir le grade de docteur de l'Ecole Nationale Supérieure des Télécommunications spécialité Informatique et Réseaux. La thèse « *Partage de charge et Ingénierie de Trafic dans les réseaux MPLS* » a été réalisée par *M. Ramon Casellas* <casellas@infres.enst.fr> et dirigée par *M. Daniel Kofman* <kofman@infres.enst.fr> . Cette thèse a été financée à l'aide d'une bourse allouée par France Télécom, division Recherche et Développement (FTRD/DAC/ISIS, responsable *M. James Roberts*), contrat NUM 991B174.

Les travaux de recherche de cette thèse ont eu lieu à l'Ecole Nationale Supérieure des Télécommunications, 46, rue Barrault 75634 Paris Cedex 13.

Contexte

Le contexte de cette thèse est constitué par un ensemble de projets développés à l'ENST et à France Télécom Recherche et Développement ayant pour objectif la conception d'un réseau haut débit multi-service. Dans la conception d'un tel réseau on peut découpler (au moins jusqu'à un certain degré) les différents domaines de recherche et développement : au niveau du transport, au niveau de la définition des services, au niveau de l'adaptation des applications pour bien tirer parti des services offerts par le réseau. Au niveau de transport, la conception et le déploiement d'un réseau multi-services doivent absolument prendre en compte le fait que différents services ont des besoins de QoS différents. Dans cette perspective, le routage sensible à la QoS, (ou de manière générique routage contraint) et l'ingénierie de trafic jouent des rôles très importants dans le dimensionnement, la conception, la mise en place et l'optimisation d'un réseau performant.

Structure du document

Cette thèse est structurée de la façon suivante : Après le chapitre 1, *Introduction Générale*, la première partie illustre le cadre architectural et technologique et soulève les problèmes de l'ingénierie de trafic : le chapitre 2 fait une synthèse de l'architecture MPLS et le chapitre 3 identifie les extensions de cette dernière en cours de normalisation. Nous donnons une introduction succincte à l'ingénierie de trafic dans le cadre des réseaux MPLS tout en présentant ses éléments les plus significatifs dans le chapitre 4.

La deuxième partie, *Optimisation et Dimensionnement*, commence au chapitre 5, *Outils Mathématiques*, qui définit la notion de bande passante effective, son rôle dans la modélisation du trafic et l'évaluation de performances, ainsi que les principaux résultats de cette théorie appliquée aux files d'attente. Les chapitres suivants présentent le noyau des travaux de cette thèse : un modèle nous permettant de trouver des règles de dimensionnement pour le partage de charge (chapitre 6, *Partage de charge sur une topologie multi-lien*), les extensions du modèle pour le support de capacités variables (chapitre 7, *Partage de charge à Capacité variable*), les enjeux de leur extension aux réseaux (chapitre 8, *Extensions aux Réseaux*) et finalement une approche heuristique pour l'optimisation des réseaux qui tente de généraliser et d'unifier les résultats précédents (chapitre 9, *Ingénierie de trafic*). On donne à la fin nos conclusions et perspectives. Les annexes présentent des notes de lecture concernant la théorie des grandes déviations (annexe A, *Techniques des Grandes Déviations*),

le projet *MPLS et Linux*, qui a comme objectif l'implémentation des plans d'usager et de contrôle MPLS et leur intégration au noyau du système d'exploitation Linux et quelques éléments logiciel que nous avons utilisés.

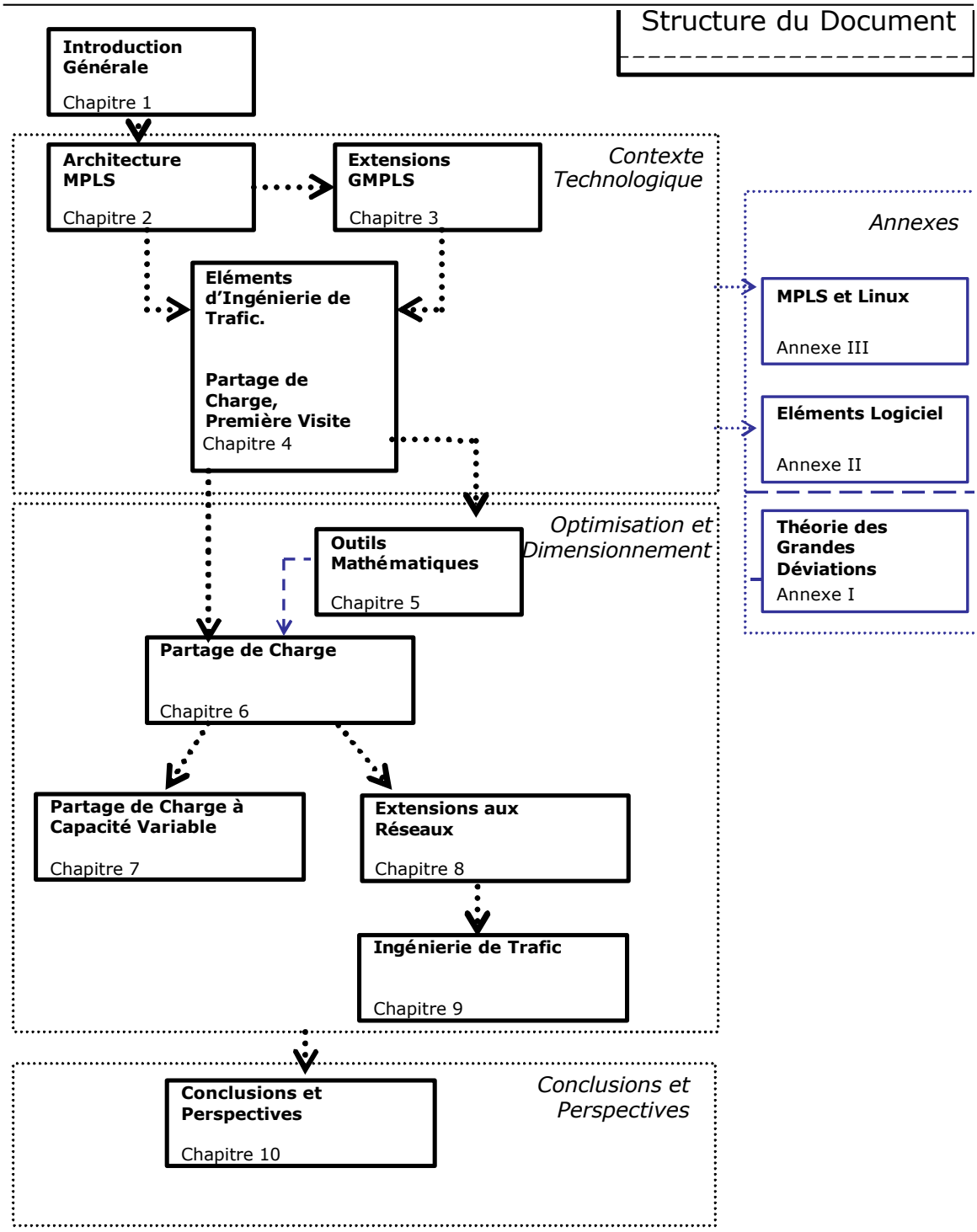


Fig. 0.1: Structure du Document

Table des matières

1. Introduction Générale

L'énorme succès de l'Internet provient principalement de deux facteurs : la simplicité d'accès au réseau et la disponibilité de services adaptés aux besoins des utilisateurs. A l'origine, le protocole IP avait été conçu comme un protocole permettant l'interconnexion des réseaux hétérogènes *(IP over everything)* pour le transport de données de bout en bout. Cela avait permis aux développeurs de services de faire abstraction de la technologie de transport et de développer, sur des API ouvertes [1], la première génération de services : le courrier électronique (e-mail), le transfert de fichiers (FTP) ou l'accès distant (telnet, rlogin). Ces services n'imposaient pas de contraintes de QoS fortes et l'architecture IP ne les avait pas prises en compte. Par ailleurs, l'usage «non commercial» avec un faible trafic global ne justifiait pas la mise en place d'une ingénierie de trafic. L'ingénierie de trafic, ou optimisation des réseaux opérationnels, peut être résumée de façon simpliste par la phrase *put the traffic where the resources are,* par opposition à *put the resources where the traffic demand is* qui pourrait définir la planification des réseaux (bien que ces deux notions ne s'appliquent pas à la même échelle de temps).

La deuxième génération de services se centre sur le développement du Web, l'ouverture commerciale du réseau et la croissance du trafic qui s'ensuit. Pourtant, jusqu'à un certain niveau, le réseau s'avère raisonnablement gérable : l'ingénieur réseau connaît l'emplacement des serveurs importants, il peut donc identifier les points de concentration du trafic et dimensionner et planifier le réseau en conséquence.

La troisième génération de services, par contre, impose des contraintes bien plus fortes. En effet, il y a d'une part l'avènement de services de données du type «peer-to-peer» qui rendent le trafic beaucoup plus imprévisible. Nous ne faisons pas référence ici aux phénomènes de dépendance longue connus depuis longtemps, mais à une très forte non stationnarité du trafic, provenant du fait que l'information est beaucoup plus distribuée que sur le Web. D'autre part, on passe au «everything over IP» qui s'ajoute au «IP over everything». En effet, on observe une tendance vers des réseaux multi-services basés sur IP (Téléphonie sur IP, Next Generation Networks, Web Services, UMTS release 5).

Le réseau se doit donc d'offrir des modèles de service évolués et son utilisation en tant que réseau d'opérateur multi-services impose le développement d'une ingénierie de trafic adaptée à cet usage. En effet, l'objectif d'un opérateur consiste à trouver un compromis entre (a) une gestion optimale de ses ressources, (b) une minimisation des coûts associés à l'exploitation et à la gestion du réseau, et (c) la possibilité d'offrir un large éventail de services à ses clients.

L'architecture *Multi Protocol Label Switching* (MPLS), en cours de normalisation par l'IETF, s'avère un outil intéressant pour permettre une ingénierie de trafic évoluée dans le contexte des réseaux de paquets tout en évitant l'approche historique IP sur ATM qui nécessite deux plans de contrôle. Néanmoins, de nombreux problèmes liés à l'Ingénierie de Trafic doivent être résolus avant qu'un

[1] Citons, par exemple, la BSD sockets library

réseau MPLS ne puisse fournir de manière efficace des service hétérogènes opérationnels.

Dans cette thèse nous proposons des mécanismes d'ingénierie de trafic applicables dans le cadre d'un réseau d'opérateur. Le contexte technologique est fourni par l'architecture et la technologie MPLS, mais certains résultats sont suffisamment génériques pouvant être appliqués à d'autres contextes. L'objectif de nos travaux est double : tout d'abord, évaluer les performance des réseaux autorisant le partage de charge en proposant des modèles mathématiques adéquats, puis proposer des règles simples pour le dimensionnement de ces réseaux.

Plus précisément, nous nous intéressons dans une première étape à la problématique du partage de charge entre différents chemins reliant deux points. La première partie de l'étude porte sur la topologie multi-lien. La possibilité qu'offre MPLS d'introduire quand c'est nécessaire une approche «orientée connexion» lève la contrainte de «partage de charge uniquement entre chemins de coûts égaux», caractéristique des protocoles de routage IP classiques comme *Equal Cost Multipath* (ECMP). Cela nous permet de prendre en compte des chemins de longueur arbitraire afin de répondre plus efficacement au problème de la sporadicité du trafic. Nous obtenons des conditions d'optimalité du partage de charge et nous déterminons, sous certaines conditions, des règles simples de dimensionnement.

La deuxième partie de l'étude étend le modèle précédent au partage de charge de bout en bout : nous proposons un modèle qui prend en compte la nature non déterministe de la capacité de chaque chemin vue par un observateur placé à un point d'entrée du réseau. La variabilité temporelle de la capacité de transmission d'un tunnel MPLS est prise en compte dans les problèmes d'optimisation à l'aide de la notion de *«capacité effective».* Nous montrons l'insuffisance des règles de dimensionnement se basant sur des moyennes et proposons des solutions plus précises.

L'extension des travaux précédents à un réseau s'avère complexe, principalement du fait de la difficulté de caractériser les modifications des propriétés stochastiques des flots de trafic lors de leur passage par un lien de transmission et de l'effet tronquant et lissant des files d'attente. Néanmoins, nous proposons une approche pragmatique sous l'hypothèse de «petits buffers» nous permettant une extension aux réseaux. Nous formulons un problème d'optimisation, analysons différentes topologies et faisons le lien avec des problèmes de théorie des graphes et d'optimisation combinatoire.

Nos travaux s'appuient sur la théorie des bandes passantes effectives et sur la théorie de grandes déviations appliquée aux files d'attente. La notion de bande passante effective correspond à un changement d'échelle de la transformée de Log Laplace du processus stochastique représentant le travail produit par une source sous les hypothèses assez génériques de stationnarité et d'ergodicité. Cette notion caractérise les propriétés statistiques de la source, telles que sa sporadicité. Cette sporadicité est prise en compte lors de la formulation que nous faisons des problèmes d'optimisation. Les fonctions objectives que nous proposons reflètent des critères réels de QoS et leur calcul utilise des résultats récents de la théorie des grandes déviations. D'une manière générale, la théorie des grandes déviations nous permet d'obtenir des estimateurs probabilistes de certaines métriques intéressantes, telles que le taux de perte ou le délai, sous un régime asymptotique déterminé (c.-à.-d. quand un paramètre du système est *assez grand*). Citons, notamment, le résultat connu comme *l'asymptotique de grand nombre d'usagers,* qui donne des équivalents logarithmiques et/ou des expressions asymptotiquement correctes de la queue de distribution stationnaire du travail cumulé dans une file d'attente ainsi que du taux de pertes quand le nombre de sources multiplexées est suffisamment grand, hypothèse largement vérifiée dans le contexte d'un coeur de réseau d'opérateur.

Cet outil mathématique s'est avéré particulièrement efficace pour analyser les problèmes de dimensionnement cités et pour déduire des règles d'ingénierie que nous considérons utiles dans un contexte de réseau d'opérateur.

Introduction

The huge Internet growth of the past years is mainly due to the sum of two factors : the simplification of network access for users and the availability of a wide range of services adapted to the users' needs.

At the beginning, the IP (Internet Protocol) protocol was conceived in order to interconnect several heterogeneous networks *(IP over everything)*, thus allowing end-to-end data transfers. Service implementors could abstract from a particular transport technology and develop, using open Application Programming Interfaces (APIs) \footnote{For example, the BSD sockets library} the first generation of services : e-mail, file transfer or remote login. These services were not particularly demanding in terms of quality of service, and the «non commercial» usage of the network did not justify traffic engineering. Traffic engineering, or optimization of operational networks, may be simply put as *put the traffic where the resources are,* compared to *put the resources where the traffic demand is* which applies to the notion of network planning (although these terms do not usually operate in the same time scale).

The second generation of services is given by the development of the World Wide Web, the deployment of commercial services and the increase of network traffic. However, to some extent, the network is still easy to manage : network operators know the placement of the most demanded servers and can identify traffic concentration points and dimension / plan the network according to these factors.

The third generation of services, however, adds a new set of constraints. Indeed, new data services like «peer-to-peer» networks make network traffic much more difficult to predict. Moreover, the network evolves to the «everything over IP», paradigm, on top of the aforementioned «IP over everything» : there is steady and clear evolution towards multi-service networks based on the IP protocol (Voice over IP, Next Generation Networks, Web Services, UMTS release 5).

Today's evolved data service models require the deployment of an adapted traffic engineering. Network operators must find a trade-off between (a) an optimal use of network resources, (b) a reduction of management costs (OPEX, CAPEX) and (c) the possibility of offering a wide range of services to customers.

The *Multi Protocol Label Switching* (MPLS) architecture, provides an excellent framework for an advanced traffic engineering, without the limitations of having two unrelated control planes like in legacy architectures (as in IP over ATM). Nevertheless, some issues and open problems remain and must be solved before being able to provide advanced data services.

In this dissertation, we propose traffic engineering mechanisms in a carrier's network. Although these mechanisms are conceived in the framework of MPLS networks, they also apply to other contexts. Our objective is twofold : first, evaluate their performance in load sharing enabled networks by proposing adequate mathematical models and second, obtain simple engineering rules and guidelines for an adequate network dimensioning.

More concretely, we first consider the load sharing problem between two network endpoints. In a first step, we focus on the ”multi-link” topology : two network devices are connected by a given number of links. Since the MPLS framework adds a ”connection oriented layer” to IP networks, it is possible to implement load sharing schemes that are not constrained by the «load sharing between equal cost paths only» property which characterizes legacy IP routing protocols like *Equal Cost Multi-path* (ECMP). In this sense, it is possible to use arbitrarily length paths in order to better respond to the problem of traffic burstiness. We obtain optimality conditions and we determine, under certain hypothesis, simple dimensioning rules.

The second part extends the previous model by taking into account the non deterministic nature of the end to end capacity of the paths as seen by a user placer an the network ingress. The time varying property is included in the revised optimization problem by using the notion of *«effective capacity».* We show that some rules based on average values may not be appropriate.

The extension of the obtained results to a network is not simple, since it is difficult to characterize the changes of the stochastic properties of flows when they are multiplexed in a queue modelling a transmission link. Nevertheless, we propose a pragmatic approach for traffic engineering under the assumption of «small buffers». We formulate a new optimization problem, we analyze different topologies and we establish the link with graph theory and classical combinatorial optimization problems.

Our work is based on the theory of effective bandwidths and the large deviation theory. The notion of effective bandwidth corresponds to the scaled logarithmic moment generating function of the stochastic process modelling the work produced by a data source, under the quite realistic hypothesis of stationarity and ergodicity. This characterization takes into account the traffic burstiness, and appears in the different optimization problems that are proposed. The objective cost functions that we formulate map directly to real world QoS criteria, re-using recent results from the large deviation theory. In general terms, this results concern the estimation of interesting metrics, like asymptotic overflow probabilities and loss rates.

This framework appears well adapted to our objectives : the analysis of the aforementioned dimensioning problems and the deduction of engineering rules of interest in core notworks.

Première partie .

Contexte Technologique

2. L'architecture MPLS

2.1. Introduction et Motivation

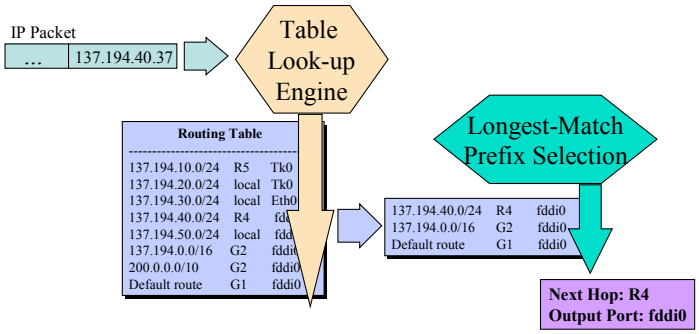
Le but de ce chapitre et du suivant est de présenter les principaux éléments de l'architecture *Multi Protocol Label Switching,* (MPLS) que l'on peut traduire par « commutation d'étiquettes multi protocolaire » ainsi que ses extensions connues sous le terme GMPLS (MPLS Généralisé). Cette architecture constitue le contexte technologique de cette thèse.

Le chapitre est structuré de la façon suivante : après la présente section, qui présente succinctement l'architecture et son évolution historique, nous présentons dans la section 2.2 les éléments et les entités importants, la terminologie utilisée et les mécanismes et protocoles de distribution d'étiquettes. La section 2.3 énumère les principaux objectifs de l'architecture, et la section 2.4 présente les conclusions.

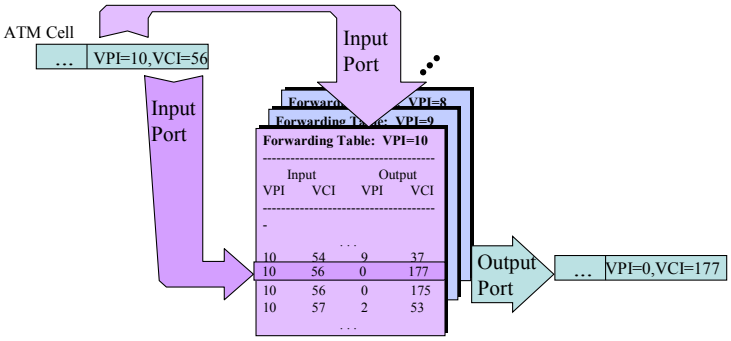
L'architecture MPLS a été développée autour du paradigme de *commutation d'étiquettes* : les informations nécessaires pour acheminer (aiguiller) [1] une unité de données sont obtenues à l'aide d'une valeur (dont l'appellation générique est étiquette) que nous allons, pour l'instant, supposer être codée dans l'unité elle-même. Ce paradigme, existant déjà dans le contexte des réseaux Frame Relay (l'étiquette = DLCI) ou ATM (VPI/VCI), s'applique principalement au plan d'usager MPLS *(forwarding user plane).* Nous allons voir dans la suite comment les premiers efforts autour de la technologie MPLS essaient d'adapter ce paradigme au monde IP, en substituant l'acheminement IP classique par un acheminement basé sur une valeur de taille fixe, permettant une amélioration des performances des équipements réseaux. A ce propos, la figure 2.1 (a) illustre l'acheminement IP classique : suite à la réception d'un paquet IP, le routeur parcourt la table construite à partir des informations topologiques et choisi l'entrée vérifiée par le paquet ayant le masque le plus longue. La figure 2.1 (b) illustre l'acheminement de cellules ATM, qui utilise le paradigme de la commutation d'étiquettes (commutation de cellules) : le port et les VPI/VCI de sortie sont choisis à partir d'une table à l'aide du port et des VPI/VCI d'entrée. Le plan de contrôle, qui définit l'ensemble de procédures et protocoles de signalisation pour l'établissement de tunnels MPLS nommés *Label Switched Paths ou LSPs,* est séparé formellement du plan d'usager. Plusieurs auteurs sont d'accord pour voir de façon simpliste MPLS comme une couche permettant une meilleure intégration des couches 2 et 3. Le qualificatif de « multi protocolaire » est justifié par son caractère a priori agnostique en ce qui concerne le protocole de niveau 3, même si le protocole IP est clairement l'acteur dominant et influence d'une certaine façon son développement, et en ce qui concerne le niveau 2. Autrement dit, l'architecture ne fait pas d'hypothèses limitant la technologie de transport, à quelques normes près spécifiant l'implémentation pour une technologie de niveau 2 comme ATM ou FR. Cet agnosticisme (au moins théorique) par rapport au protocole de niveau 3 est dû à la séparation du plan de contrôle

[1] Il est important de faire la différence entre les fonctions de routage *(ang. routing)* et d'acheminement *(ang. forwarding).* Le routage concerne la découverte des routeurs voisins et l'échange d'informations topologiques grâce à un protocole de routage, et l'acheminement correspond à la fonction d'aiguillage d'un paquet reçu sur un port d'entrée vers un port de sortie selon une table d'acheminement construite une fois que le routeur possède une connaissance de la topologie du réseau.

et du plan usager et il facilite la conception de solutions où MPLS elle-même agit comme technologie de transport commune pour des protocoles différents. Citons un exemple qui sera présenté en détail ultérieurement : les réseaux privés virtuels, dans lesquels MPLS agit comme technologie de transport pour le protocole IP, mais aussi pour le transport transparent de cellules ATM, permettant ainsi l'offre de services VPN de niveaux 2 et de niveaux 3, sans le coût associé à la gestion de réseaux différents.



(a)



(b)

Fig. 2.1: Acheminement dans les réseaux IP et dans les réseaux ATM

Avant de décrire les éléments de l'architecture MPLS, on doit considérer les architectures ayant fortement influencé son développement et sa normalisation postérieure. Parmi celles-ci, deux solutions propriétaires sont notables : Cisco Tag Switching et Ipsilon IP switching [48]. Ces deux solutions présentent quelques notions clés : d'abord, un équipement mixte, intégrant une matrice de commutation ATM et une partie de contrôle IP ; ensuite, une vraie séparation des plans de contrôle et usager. Le routage IP dynamique contrôle l'établissement de circuits virtuels, à la place des méthodes comme l'utilisation de logiciels d'administration, la configuration manuelle, ou même le protocole de routage PNNI.

2.1.1. Evolution Historique

Identifier et supprimer les « goulots d'étranglement » était et reste un problème classique pour l'ingénieur réseau. Il paraît raisonnable d'affirmer qu'au cours de l'évolution des réseaux, ces goulots d'étranglement ont oscillé entre les domaines de la transmission et de la commutation. D'une part, on constate le développement de technologies comme le multiplexage de longueur d'onde (DWDM), qui apparaissent avec la promesse d'une capacité presque illimitée et d'autre part, les constructeurs d'équipements présentent des produits de plus en plus performants capables « a priori » d'acheminer à la vitesse du lien de transmission sans perte de performances (« line rate forwarding ») tout en assurant un ensemble de fonctionnalités plus évolué que le simple acheminement (par exemple le filtrage).

Ainsi, depuis 1995-1996 environ, les besoins de commutation des routeurs IP augmentent sans cesse, suite à la croissance du trafic Internet. Les premières architectures des routeurs et les algorithmes d'acheminement IP présentaient une approche naïve du problème : les tables de routage étaient parcourues de façon linéaire, les entrées vérifiant le préfixe d'adresse étaient stockées dans une zone mémoire temporaire, et celle ayant un masque de longueur maximale indiquait l'adresse du routeur suivant. La communauté s'est vite rendu compte de la nécessite de concevoir des solutions plus performantes afin de répondre rapidement à la croissance du trafic Internet.

Une première réponse au problème du transport IP fut donnée par la technologie ATM *(Asynchronous Transfer Mode)*. ATM définit un mode de transfert asynchrone pour un réseau à intégration de services à large bande et a été conçu pour prendre en charge une qualité de service native et des interfaces normalisées à haut débit. En proposant des solutions permettant le transport d'IP sur ATM, il est possible de réutiliser le savoir-faire et la technologie (par exemple, les matrices de commutation) développés dans un contexte purement ATM. Ainsi, la superposition en couches IP sur ATM sur SDH devint courante, et un certain nombre d'architectures et de protocoles virent le jour : *Classical IP and ARP over ATM (CLIP), Next Hop Resolution Protocol (NHRP), MultiProtocol over ATM (MPOA)* entre autres (voir p.ex. [52] ou [56]). On peut néanmoins se poser la question de la nécessité de la couche SDH, d'autant plus que les normes ATM définissaient la possibilité d'envoyer des cellules sur la couche physique (*cell based*). A ce propos, SDH était une technologie maîtrisée, largement déployée[2], fournissant l'horloge pour la synchronisation et des mécanismes de protection et de restauration.

Cette superposition en couches présente néanmoins des limitations : une efficacité protocolaire qui pouvait s'avérer insatisfaisante et la nécessité de gérer plusieurs plans de contrôle et de gestion. La couche ATM au cœur du réseau est justifiée dans une grande mesure par sa fonctionnalité de multiplexage, permettant le support des services natifs ATM et en tant que technologie de transport pour IP et Rélai de Trames (*Frame Relay*) ainsi que par les mécanismes de qualité de service disponibles, même si IP n'en profitait pas : IP se développait dans un contexte «best-effort», avec des applicatifs pas trop gourmands en qualité de service et n'était pas une source de revenu importante.

A fur et à mesure que le protocole IP devenait l'acteur dominant, deux grandes lignes de développement parallèles apparaissaient : d'un coté, l'utilisation de SONET/SDH comme technologie de transport IP (POS ou Packet Over SONET/SDH), utilisant une version simplifiée du protocole PPP (Point-to-Point Protocol) et visant à supprimer la couche ATM en cœur de réseau. PPP fournit le

[2] notamment pour les services de voix

multiplexage (rappelons, par exemple, qu'un certain nombre de protocoles comme ICMP ou ARP sont indissociables d'IP) et assure les mécanismes de détection de trames et de détection d'erreurs. Cette transition est motivée par les coûts associés à la gestion des différents plans de contrôle et de gestion, et par une efficacité protocolaire d'environ 90%, inefficace dans les liens transocéaniques. De l'autre coté, la conception de solutions visant à une meilleure intégration des couches 2 et 3, qui a donné naissance à l'architecture MPLS. Si l'on se pose la question de la raison d'être *à l'époque* de l'architecture MPLS (ou plutôt, des architectures propriétaires qui l'ont fortement influencée) la réponse est immédiate : le besoin de concevoir des dispositifs de plus en plus performants, facilitant une meilleure intégration IP et ATM, avec un taux de transfert[3] de plus en plus élevé pour répondre à l'augmentation du trafic et sans la complexité inhérente aux architectures CLIP ou MPOA.

L'approche « classique » pour le transport d'IP sur ATM utilise la couche d'adaptation AAL5 et nécessite que les équipements réseaux récupèrent la charge utile de la couche AAL5 pour consulter l'entête du paquet et décider du routeur suivant vers lequel acheminer le paquet, procédure coûteuse qui doit être réalisée à chaque saut. A ce propos, plusieurs architectures propriétaires et concurrentes sont apparues, avec des promesses concernant la performance de leurs équipements. *L'idée géniale* peut se résumer dans les points suivants :

1. L'utilisation d'équipements mixtes, avec les matrices de commutation ATM sous plan de contrôle IP. Le rôle de la technologie ATM est réduit à un rôle de transport.
2. L'association d'une valeur de taille fixe à des flots ou à des agrégats de trafic. Cette valeur est codée dans les champs VPI/VCI.
3. L'acheminement IP est réalisé en consultant l'étiquette, et non en consultant l'entête IP. On est désormais capable d'acheminer sans besoin de récuperer la charge utile de la couche AAL5. Ceci permet une amélioration de la performance de l'acheminement IP classique.
4. La définition d'une composante du plan de contrôle IP qui s'occupe d'associer une valeur à des flots ou à des agrégats IP et de les distribuer aux équipements concernés.
5. L'adressage et les composantes du plan de contrôle ATM comme la signalisation et le routage deviennent superflus.

Ipsilon IP Switching

L'architecture propriétaire IP Switching utilisait la détection au vol des flots IP et une procédure de signalisation se déclenchait pour l'établissement, la gestion et le relâchement des connexions. La relation connexion ATM - flot IP était gérée par un plan de contrôle et utilisait un protocole de distribution propriétaire. Cette solution était pénalisée en cœur du réseau par la difficulté de gérer un état par flot.

Cisco Tag Swithing

Tag Switching associait à chaque entrée ou préfixe de la table de routage (nous verrons comment cette idée a été généralisée avec la notion de FEC) une valeur entière *(Tag)* de signification purement local entre deux dispositifs voisins (physiques ou logiques) identifiant sans ambiguïté un préfixe d'adresse :

[3] le taux de transfert pouvant s'exprimer en paquets par seconde ou pps

ainsi, un routeur susceptible de recevoir du trafic communiquait ses tags à ses voisins, grâce à des protocoles de distribution adaptés. La construction de tables de correspondance entre tags et préfixes permettait d'améliorer les performances des routeurs en réalisant un acheminement des paquets basé sur la commutation de tags.

La raison d'être de MPLS, aujourd'hui

La raison d'être de MPLS à évolué. Nous avons vu comment MPLS (alors en cours de normalisation par l'IETF, en reprenant les solutions propriétaires) fournissait une réponse au problème de la croissance du trafic et au besoin d'améliorer les performances des routeurs avec une approche mixte IP-ATM où la technologie ATM (ou plutôt les matrices de commutation ayant été développées dans le contexte de cette technologie) seraient réutilisées pour l'acheminement des paquets IP. Néanmoins, le développement en parallèle de circuits spécifiques (ASICs), la conception d'algorithmes optimisés de recherche dans les tables de routage et l'apparition de processeurs performants dédiés permettant de plus en plus de fonctionnalités sans perte de performance rendaient difficile la justification de MPLS si son seul intérêt était une amélioration marginale des performances.

Néanmoins, la communauté s'est vite rendu compte que les atouts de MPLS sont plus nombreux. Tout d'abord, l'ajout grâce a cette technologie d'une couche orientée connexion aux réseaux IP ouvre la porte à l'introduction de la qualité de service et de l'ingénierie de trafic IP et à la réutilisation des études et travaux existants adaptés aux réseaux à commutation de paquets orientés connexion. Ensuite, l'architecture MPLS peut être déployée de façon progressive et permet la coopération et la cohabitation des technologies classiques ATM et IP. MPLS est devenue un fournisseur de mécanismes d'ingénierie de trafic, dont les éléments les plus importants seront présentés au chapitre 4, et qui permet une optimisation des réseaux tout en assurant un certain niveau de qualité de service (mesurable en termes de bande passante minimale, de délai ou de métriques administratives). Enfin, l'intégration de différentes architectures de réseaux privés virtuels (VPN) sous une infrastructure unique et l'unification du plan de contrôle simplifient les coûts de gestion et d'opération.

2.2. Eléments de l'architecture MPLS

2.2.1. Label Switch Router

D'une manière générique on appelle *Label Switch Router* ou LSR tout routeur (dispositif ou élément réseau) faisant partie du domaine MPLS. Plus particulièrement, un LSR est capable d'acheminer des paquets dans le plan usager en utilisant le paradigme de commutation d'étiquettes et d'interpréter les protocoles du plan de contrôle associés. [4]

2.2.2. Classe d'Equivalence pour l'Acheminement

De façon abstraite, l'ensemble de tous les paquets appartenant à un protocole de niveau 3 peut être vu comme un «espace», sur lequel on peut définir une partition étant donné un ensemble de propriétés. Cette partition va définir des classes d'équivalence. Ainsi, la notion de Classe d'Equivalence pour

[4] Bien sûr, vis-à-vis des produits disponibles sur le marché, il y a toujours la problématique de la *conformité* aux différents protocoles et normes.

l'Acheminement *(ang. Forwarding Equivalence Class ou FEC)* correspond à chaque classe résultante d'une partition du trafic en ce qui concerne l'acheminement, et implique que tous les paquets appartenant à une même classe seront traités, à niveau de l'acheminement, de façon identique [5]. Même si la notion de FEC apporte une généralité et une granularité importante, aujourd'hui, la FEC la plus utilisée est la FEC *préfixe d'adresse destination* : deux paquets appartiennent à la même FEC si leur préfixe d'adresse destination est identique. Ex : on parle de la FEC *137.194/16* pour désigner l'ensemble de paquets destinés à l'ENST. Remarquons que la notion de FEC est à la base de tous les mécanismes concernant la répartition de trafic.

Formalisation de l'acheminement IP classique : en utilisant la notion de FEC, l'acheminement IP classique peut être vu comme la succession de deux tâches : (a) faire correspondre (classer) le paquet entrant à une FEC, et (b) faire correspondre la FEC au routeur suivant. Ces deux tâches se répètent à chaque saut. Nous allons voir comment MPLS réalise le classement une seule fois, à l'entrée du domaine.

Relation FEC-Etiquette : nne *étiquette (label)* est une représentation non ambiguë d'une FEC entre deux LSR. Pour un transfert de données unidirectionnel entre les routeurs R_u et R_d, on nomme R_u le routeur *Upstream (amont)* et R_d le routeur *Downstream (aval)*. Le terme *distribution d'étiquettes* désigne toute procédure destinée à associer FEC et Etiquette globalement dans le réseau.

2.2.3. Le paradigme de la commutation d'étiquettes

Nous avons déjà évoqué le paradigme de la commutation d'étiquettes. En utilisant une terminologie MPLS, le mécanisme de base de MPLS est le suivant [6] : au niveau du plan usager, le paquet IP entre dans le *domaine* MLPS via le routeur d'entrée *(Ingress LSR ou I-LSR)*. Le I-LSR classe le paquet dans une FEC, selon des critères tels que les adresses source et destination. Le I-LSR consulte la table *FTN (FEC to NHLFE, Next Hop Label Forwarding Entry)*. Cette table pointe vers une deuxième table, NHLF, qui indique le routeur suivant et l'opération à réaliser, c'est à dire, l'ajout d'une étiquette *(Label Push)*et sa valeur. **Remarquons la base de l'ingénierie de trafic MPLS : un routeur détermine le routeur suivant (le «next hop») à partir de la table NHLF, et celui-ci peut être différent de celui imposé par le protocole de routage.** L'acheminement dans le domaine MPLS est basé sur l'étiquette (et donc, les LSRs ne regardent pas l'entête IP) en utilisant le paradigme de la commutation d'étiquettes *(label swap / switching)* et une table nommé *ILM (Incoming Label Map)*. Finalement, le paquet étiqueté est reçu par le routeur de sortie du domaine *(Egress LSR ou E-LSR)* qui enlève l'étiquette *(label pop)* et procède à un acheminement classique. La figure 2.2 (a) illustre ce paradigme dans un domaine MPLS «plat».

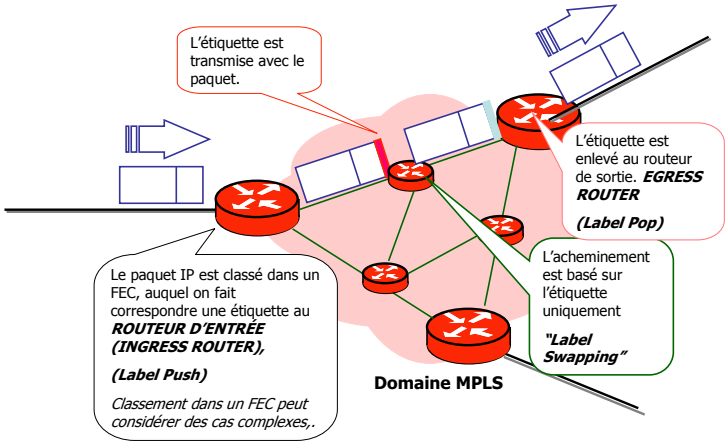
Ce paradigme peut être généralisé à un contexte «hiérarchique» (cf. figure 2.2 (b)) sous les conditions suivantes :

1. La notion d'étiquette est généralisée à *une pile d'étiquettes (Label Stack)*.
2. Le traitement d'un paquet IP est réalisé à l'aide de l'étiquette de plus haut niveau.
3. Les rôles I-LSR et E-LSR sont maintenant conditionnés par le niveau hiérarchique concerné.

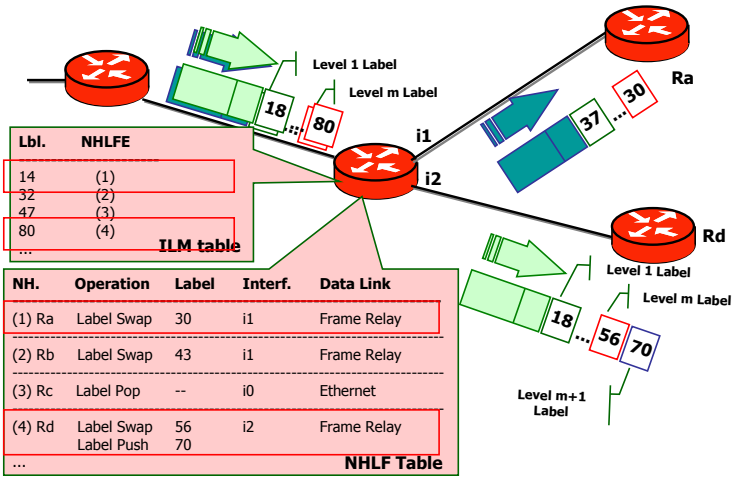
[5] Autrement dit, au niveau du point de vue de l'acheminement, deux paquets appartenant à la même classe sont identiques.

[6] Pour des raisons de clarté, nous centrerons l'exposé sur IP comme protocole de transport, mais le paradigme reste générique.

La section 2.2.5 décrit les bénéfices d'une telle hiérarchie.



(a)



(b)

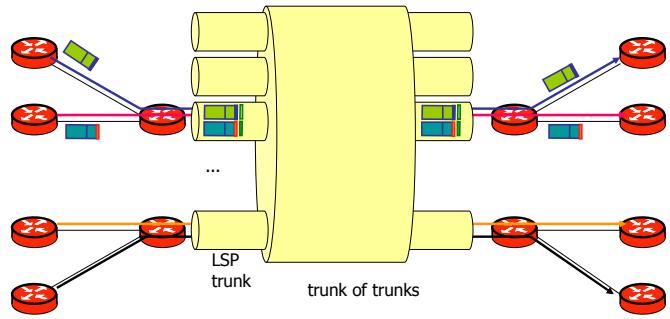
Fig. 2.2: Mécanisme de Base MPLS : Commutation d'étiquettes

2.2.4. Label Switched Path

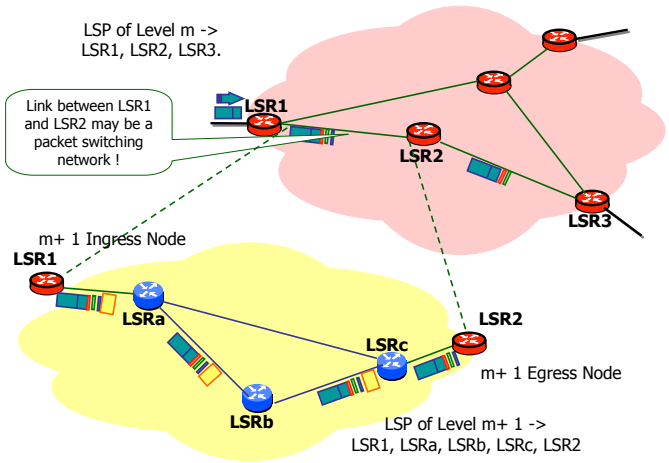
Remarquons que le fait de classer un paquet IP dans un FEC donné détermine son étiquette de sortie, le routeur suivant et son traitement ultérieur dans le domaine. D'une façon informelle, ce classement détermine la liste de routeurs (ou de façon équivalente, de liens) à traverser, et donc le chemin. C'est ainsi que l'on définit la notion de *Chemin à Commutation d'Etiquettes (ang. Label Switched Path ou LSP)* . On trouve parfois le terme «tunnel MPLS».

2.2.5. La hiérarchie MPLS

La hiérarchie MPLS généralise celle de la technologie ATM à deux niveaux (VPI/VCI). Les avantages de la hiérarchie MPLS sont les suivants : (a) *l'agrégation de trafic sur plusieurs niveaux,* (b) *la possibilité de commuter les étiquettes sur différents niveaux hiérarchiques, en assurant un acheminement de bout en bout transparent pour un niveau donné.* La figure 2.3 illustre ces notions.



(a)



(b)

Fig. 2.3: Hiérarchie MPLS, Applications

Détail d'encapsulation : l'architecture MPLS fait l'hypothèse que l'étiquette apparaît explicitement codée dans le paquet (nous verrons au chapitre suivant comment GMPLS lève cette contrainte). Deux possibilités existent : la première, consiste à tirer parti de certaines particularités de la technologie de transport sous-jacente pour y coder l'étiquette (citons, par exemple, l'utilisation des champs VPI ou VCI ATM ou du champ DLCI FR). L'inconvénient de cette approche, c'est que le nombre maximal de niveaux hiérarchiques est limité par la technologie. La deuxième consiste à préfixer le

paquet IP avec la pile d'étiquettes (permettant donc un nombre de niveaux hiérarchiques arbitraire [7]). Autrement dit, la pile d'étiquettes est codée entre l'entête de niveaux 2 et l'entête de niveau 3 suivant un format particulier dit *Shim Header*. L'IETF a de plus défini un ensemble de normes permettant de trouver un compromis entre ces deux possibilités, voir p.ex. [64].

2.2.6. La distribution d'étiquettes

L'architecture MPLS **ne fait aucune hypothèse concernant le protocole de distribution d'étiquettes utilisé.** Cette propriété, conséquence directe de la séparation des plans de contrôle et d'usager, est plutôt positive : vis-à-vis du plan usager, rien ne change si les tables FTN ou ILM ont été approvisionnées manuellement ou résultant d'un établissement dynamique à l'aide des protocoles de signalisation conçus à cet effet. Le premier LDP *Label Distribution Protocol* s'avère simple, mais fortement limité. Deux protocoles de signalisation plus complets (au sens où ils permettent de mettre en place des mécanismes d'ingénierie de trafic) sont en cours de normalisation par l'IETF : RSVP-TE [95] et CR-LDP [12].

2.2.6.1. LDP

Le protocole LDP *Label Distribution Protocol* est un protocole de signalisation (plus précisement, de distribution d'étiquettes) héritier du protocole propriétaire TDP *Tag Distribution Protocol.* Pour en décrire le fonctionnement, rappelons la notion de *l'arbre de plus court chemin :* pour un préfixe d'adresse, le protocole de routage classique définit implicitement un arbre de plus court chemin, arbre ayant pour racine le LSR de sortie (celui qui a annoncé le préfixe) et pour feuilles les différents routeurs d'entrée. Le routeur de sortie va *annoncer* le préfixe à ses voisins, tout y en associant une étiquette. Les messages de signalisation vont «monter» jusqu'aux routeurs d'entrée, permettant à chaque LSR intermédiaire d'associer une étiquette au préfixe. Pourtant ce protocole (par ailleurs raisonnablement simple) présente deux grandes limitations :

LSPs contraints par le protocole de routage : les LSPs établis en utilisant le protocole LDP sont contraints par le protocole de routage. Il est impossible de spécifier des routes autres que celles définies par le protocole de routage *(ang. IGP default route).*

Impossibilité de réaliser une réservation de ressources : le protocole n'a aucun moyen de spécifier des paramètres de trafic pour l'agrégat de trafic à acheminer sur le LSP.

Pour permettre la réalisation d'une ingénierie de trafic efficace, l'architecture MPLS nécessite des protocoles de signalisation et de contrôle vérifiant un certain nombre de propriétés :

1. Etablissement de LSPs sur une route explicite (liste de routeurs à traverser), non contrainte par le protocole de routage : les routes ne vont pas forcément suivre le plus court chemin imposé par le protocole de routage d'après une métrique administrative.
2. Possibilité d'une réservation (optionnelle) de ressources, au moyen d'objets d'information capables de caractériser, au moyen d'un nombre réduit de paramètres, les propriétés stochastiques du trafic. L'architecture MPLS ne spécifie pas *comment* les éléments réseaux seront capables de mettre en place de façon concrête cette réservation.

[7] Le nombre de niveaux hiérarchiques reste quand même limité par la MTU de la technologie de transport

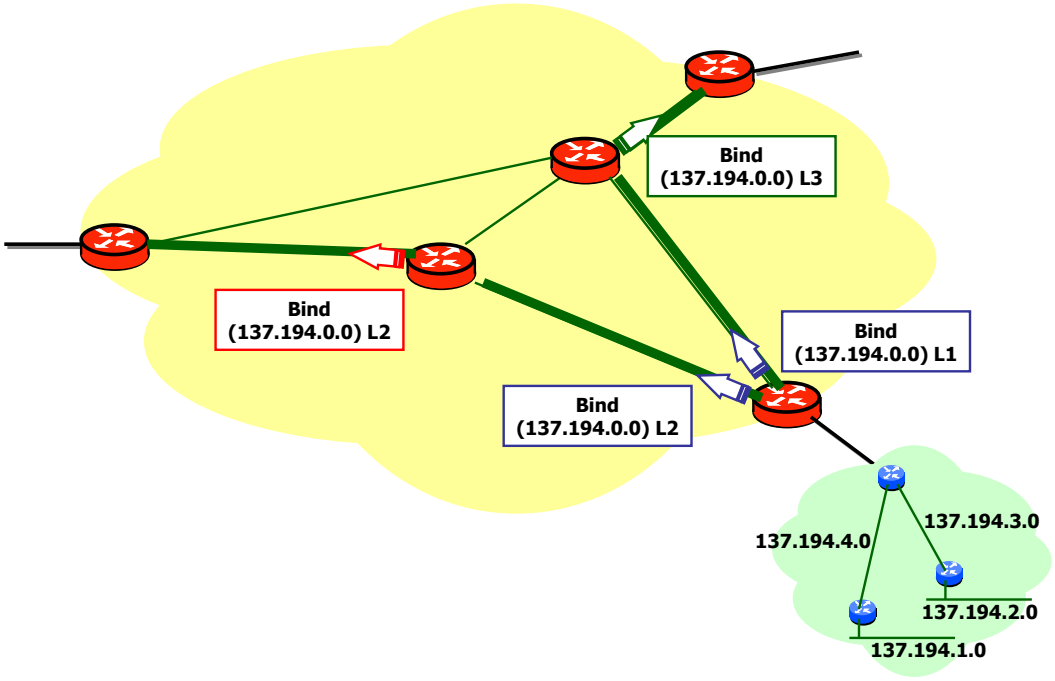


Fig. 2.4: Distribution d'étiquettes via LDP

3. Prise en charge des mécanismes de restauration basés sur MPLS *(ang. MPLS based recovery)* (voir chapitre 4, *Eléments d'Ingénierie de Trafic*), avec une gestion de priorités à l'établissement et modification de LSPs.

Deux protocoles vérifient ces propriétés (dans une certaine mesure) : RSVP-TE, et CR-LDP. En bref, ces deux protocoles ont des fonctionnalités très similaires, l'établissement d'un LSP suit une approche «aller-retour» et consiste en l'envoi et le traitement d'un message de signalisation initial qui parcourt les routeurs qui seront traversés par le LSP à établir, et le message correspondant d'acquittement et de distribution d'étiquettes.

2.2.6.2. RSVP-TE

Le protocole RSVP-TE ou *Extensions au protocole RSVP pour l'ingénierie de trafic* reprend le protocole RSVP de l'IETF, initialement conçu pour la réservation de ressources pour des micro flots (par exemple des connexions TCP entre deux points terminaux), et il ajoute les extensions nécessaires pour la prise en charge de routes explicites MPLS. Le problème du passage à l'échelle de RSVP est «résolu» en faisant une réservation de ressources entre deux routeurs (typiquement le routeur d'entrée et celui de sortie) pour *un agrégat* de trafic (pouvant être constitué de plusieurs centaines de connexions).

2.2.6.3. CR-LDP

Le protocole CR-LDP ou *Extensions pour le routage sous contraintes du Protocole de Distribution d'étiquettes, LDP* a la lourde tâche d'ajouter au protocole LDP les extensions nécessaires pour le support de routes explicites et la réservation de ressources.

Commentaires : entre 1999 et 2002, il n'était pas clair si l'un de ces deux protocoles allait s'imposer à l'autre, et cette question avait été déléguée aux constructeurs. Il apparait actuellement que RSVP-TE tend à s'imposer comme protocole de signalisation[8].

2.3. Objectifs de l'architecture MPLS

Dans cette section nous présentons les objectifs de l'architecture MPLS que nous considérons les plus importants. Les deux premiers sont issus des architectures antérieures citées précédement.

2.3.1. Intégration IP et ATM

Cette intégration est généralisable aux couches 2 et 3, grâce au caractère multi protocolaire de MPLS. La couche orientée connexion ajoutée par MPLS ouvre la porte à l'introduction d'une qualité de service et d'une ingénierie de trafic à des flots IP.

2.3.2. Améliorer les performances des routeurs IP

L'application du paradigme de commutation d'étiquettes au plan d'usager *peut* améliorer les performances des routeurs IP, même si ceci n'est pas le plus notable de ses atouts. D'une façon générique, il n'existe pas aujourd'hui une différence de performances importante entre le paradigme de la commutation d'étiquettes et un acheminement IP optimisé.

2.3.3. Séparation des plans d'usager et de contrôle

La séparation des plans d'usager et de contrôle permet le caractère multi-protocolaire de MPLS. Au niveau du plan d'usager, les informations sont acheminées sans nécessiter une connaissance du protocole transporté. C'est au niveau du plan de contrôle que cette information est nécessaire. Par exemple, dans le message d'établissement d'un LSP utilisant RSVP-TE, un des objets d'information contient le protocole de niveau 3 acheminé. Ceci permet par exemple au routeur de sortie de savoir comment acheminer l'unité d'information une fois que celle-ci quitte le domaine MPLS.

2.3.4. Unification du plan de contrôle

L'évolution historique des réseaux a eu comme résultat le déploiement de réseaux avec des technologies hétérogènes et présentant une forte superposition de couches et un surcoût protocolaire important. Un des objectifs clés de l'architecture MPLS et qui sera présenté et justifié en détail au

[8] En Août 2002, le groupe de travail de l'IETF à proposé l'abandon des travaux de normalisation de CR-LDP, et son passage à l'état de RFC informationnel.

chapitre 3, *GMPLS ou MPLS Généralisé*, est de permettre le déploiement de technologies hétérogènes mais gardant une homogénéité en ce qui concerne le plan de gestion et de contrôle afin de minimiser les coûts de gestion et d'opération du réseau.

2.3.5. Gestion de la QoS et de l'Ingénierie de Trafic

Les éléments du plan de contrôle MPLS concernant l'Ingénierie de Trafic sont exposés en détail dans le chapitre 4, *Eléments d'Ingénierie de Trafic*. D'une façon très simpliste, on peut résumer la gestion de la qualité de service et de l'ingénierie de la façon suivante : la notion de FEC fournit une granularité suffisante pour la partition du trafic et l'établissement de LSPs avec ou sans une réservation de ressources non contraint par le protocole de routage permet à l'ingénieur réseau d'optimiser son réseau en réalisant une répartition de trafic convenable, sans être limité par la technologie sous-jacente.

2.3.6. Les Réseaux Privés Virtuels MPLS (MPLS/BGP VPN)

La possibilité d'offrir des services de Réseau Privé Virtuel en utilisant la technologie MPLS a été identifiée comme un des facteurs motivant une migration progressive. L'avantage est d'avoir un seul réseau à gérer, avec la possibilité d'un seul plan de contrôle. Aujourd'hui, MPLS permet la gestion de réseaux privés virtuels de niveau 2 et 3.

Il est intéressant de noter que l'architecture MPLS est suffisamment générique pour permettre plusieurs façons de déployer le service VPN en utilisant la commutation d'étiquettes. Néanmoins, le standard «de facto» est le RFC 2547 et le RFC 2547 bis.

Le RFC 2547bis prend en charge différents protocoles de routage et envisage la possibilité de VPN inter-AS.

MPLS/BGP VPN : Architecture : deux grandes classes de LSRs dans le réseau d'opérateur sont définies. Les routeurs de la périphérie, notés PE *(ang. Provider Edge)* et les routeurs du cœur, notés P. Les sites clients sont connectés à un routeur PE, responsable de leur gestion. Les routeurs P n'ont aucune connaissance du service VPN.

MPLS/BGP VPN : Plan de Contrôle. Une instanciation d'un protocole de routage (typiquement OSPF) est exécutée par tous les LSRs du domaine MPLS. Les routeurs PE sont connectés deux-à-deux : il existe un LSP entre deux routeurs quelconques PE. Bien sûr, l'établissement de ces LSP peut être réalisé avec une réservation de ressources. Dans la configuration la plus simple, il existe un maillage complet *(ang. full mesh)* de sessions MP-iBGPv4 entre les routeurs PE. Nous verrons par la suite que ce protocole est utilisé comme protocole de signalisation.

Etant donné que l'acheminement des paquets IP dans le réseau du fournisseur de service est réalisé en utilisant MPLS, le fournisseur de service peut utiliser des adresses privées. En plus, des adresses IP n'appartenant pas au même VPN peuvent être réutilisées. Ceci est possible car l'architecture définit une nouvelle *famille d'adresses* : la famille résultant de préfixer à l'adresse IPv4 classique un identificateur de VPN (VPN Id) unique. Autrement dit, un hôte quelconque est identifié par son adresse IPv4 *et* son identificateur de VPN. L'architecture MPLS BGP/VPN utilise les extensions multi protocolaires de BGPv4 comme protocole de distribution d'étiquettes. Les extensions multi protocolaires sont nécessaires car BGPv4 est conçu pour IPv4 (donc pas de prise en charge de

l'identificateur de VPN). Le message *update* est utilisé pour *annoncer* des préfixes d'adresse **et** l'étiquette associée au préfixe [9]. Cette approche utilise seulement des FEC équivalentes à des préfixes d'adresse. Lorsque les routeurs PE autres que celui qui a initié le message reçoivent cette annonce, ils ajoutent à la table de routage virtuelle correspondante des informations telles que le routeur suivant (routeur suivant logique PE), le site qui a annoncé le préfixe, et le VPN. Ceci est illustré par la figure 2.5 (a).

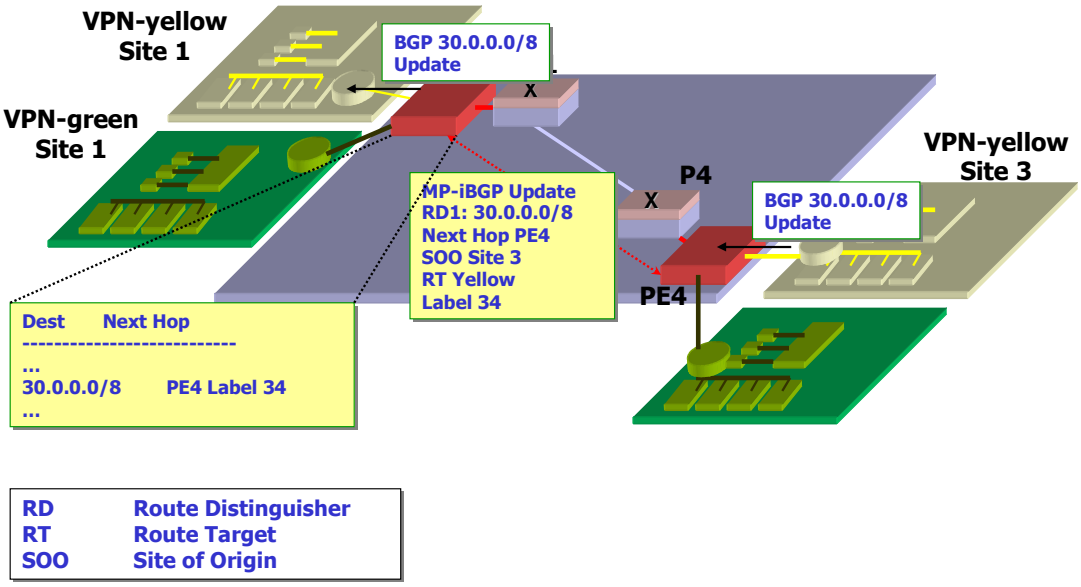
MPLS/BGP VPN : Plan usager Nous allons exposer l'acheminement de bout en bout d'un paquet IP à l'aide de la figure 2.5 (b). L'hôte source (appartenant au site 1 du VPN «vert», :10.0.0.1) envoie un paquet IPv4 à l'adresse destination (:30.0.0.1). Le paquet est reçu par le routeur PE auquel le site est rattaché (PE1), capable de reconnaître (par l'interface d'entrée, par exemple) le site origine du paquet et le VPN correspondant[10]. Avec la connaissance du VPN le routeur construit l'adresse étendue et classe le paquet dans le FEC correspondant (préfixe VPNVERT :30.0.0.0/8). Ensuite, il consulte la table de routage virtuelle (VRF) correspondante, fonction FTN (FEC to NHLFE) et il obtient l'étiquette identifiant sans ambiguïté le FEC (15) et le routeur *logique* suivant (PE4). Cette étiquette est ajoutée au paquet au niveau 1 de la hiérarchie MPLS. Etant donné que les routeurs de la périphérie (PE1 et PE4) n'ont pas de connectivité directe, l'architecture fait appel à la hiérarchie MPLS : un deuxième niveau hiérarchique MPLS permet au routeur d'entrée (PE1) d'acheminer le paquet jusqu'au routeur de sortie (PE4). La hiérarchie MPLS permet aux routeurs du backbone *n'ayant pas une connaissance du service* d'acheminer le paquet en utilisant l'étiquette externe (bleue). Lorsque le paquet arrive au routeur de sortie (PE4), celui-ci peut reconnaître la FEC à partir de l'étiquette interne (car c'est lui qui a réalisé l'association FEC - Etiquette) et acheminer le paquet de façon classique vers le site destination.

2.4. Conclusions

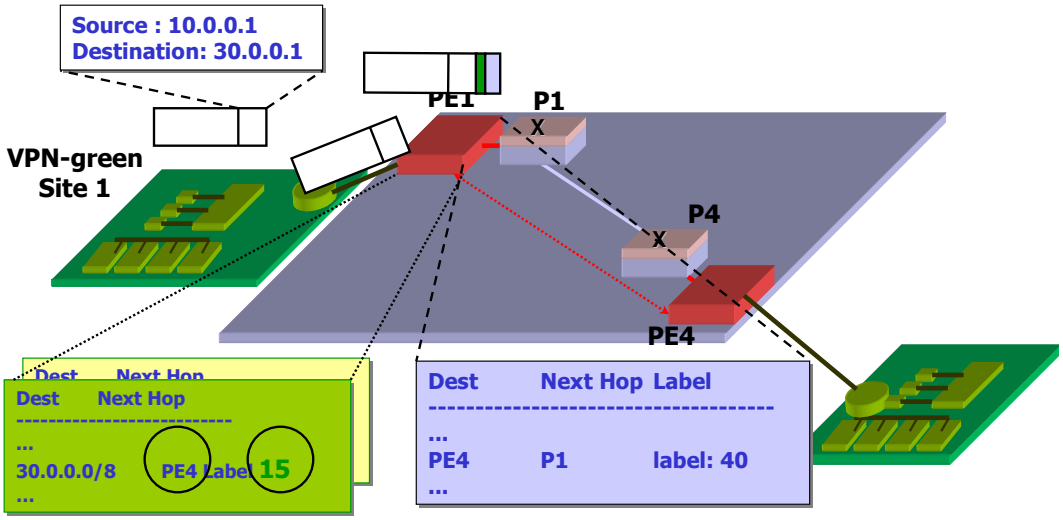
Grâce à sa nature orientée connexion, MPLS offre un cadre architectural très prometteur pour l'implémentation d'une ingénierie de trafic efficiente dans le monde IP et, en particulier, pour le partage de charge. Dans ce chapitre, nous avons présenté l'évolution historique de l'architecture MPLS, ses éléments les plus importants (LSR, LSP, FEC,..), leurs différents rôles, et les objectifs de MPLS. Dans les chapitres suivants, nous allons voir les extensions de l'architecture MPLS (GMPLS), et les composants du plan de contrôle concernant l'ingénierie de trafic.

[9] le terme souvent utilisé est *piggybacking*.

[10] Comment il fait cela, n'est pas spécifié. Si un site appartient à plusieurs VPN, le routeur PE doit pouvoir déterminer aussi le VPN, par exemple, à l'aide de l'adresse destination



(a) Plan de Contrôle



(b) Plan d'usager

Fig. 2.5: Les réseaux privés virtuels MPLS BPG/VPN

3. GMPLS ou MPLS Généralisé

3.1. Introduction

Ce chapitre est structuré de la façon suivante : la présente section illustre l'évolution historique des réseaux et justifie le besoin d'une simplification protocolaire. La section 3.2 présente une première réponse de l'IETF à ce besoin dans le contexte IP et MPLS, le *Multi Protocol Lambda Switching* MPλS, terme qui désigne une extension de MPLS visant à une meilleure intégration des domaines réseaux à commutation de paquets et réseaux optiques. L'évolution vers ce que l'on connaît aujourd'hui comme GMPLS est présentée dans la section 3.3. Cette architecture généralise la notion de *commutation,* et la section 3.4 en énumère les objectifs. La section 3.5 illustre la notion de hiérarchie GMPLS, et la section 3.6 présente les extensions architecturelles et protocolaires nécessaires à MPLS afin de concevoir un cadre architectural pouvant répondre aux besoins et aux objectives énumérés. Finalement, la section 3.7 cite les deux modèles possibles de déploiement de cette architecture et la section 3.8 conclut le chapitre.

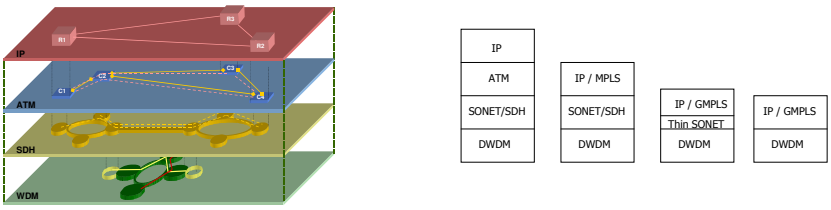
3.1.1. Evolution technologique : vers une simplification protocolaire

L'évolution historique des architectures réseaux des opérateurs est motivée d'une part par le besoin constant de répondre rapidement à la demande de nouveaux services et à la croissance du trafic Internet de leurs clients et, d'autre part, est contrainte par la nécessité d'amortir les investissements en équipements réseaux et de maintenir les services existants. Cette évolution explique le déploiement actuel des réseaux avec une importante superposition de couches protocolaires *(Overlay Networks),* superposition illustrée sur la figure 3.1. Cette superposition présente quelques inconvénients notables : (a) l'augmentation des coûts de gestion du réseau, (b) la nécessité d'une compétence technologique diverse, (c) le surcoût protocolaire et (d) la redondance indésirable de certaines fonctionnalités.

Certes, d'autres facteurs justifient l'architecture actuelle des réseaux d'opérateurs : ayant comme objectif la convergence vers un réseau multiservice, on ne peut pas négliger les avantages des différentes technologies existantes : DWDM répond au besoin de bande passante et SDH propose des mécanismes de protection et de restauration nécessaires pour les services que l'on offre aujourd'hui. Néanmoins, l'intégration de dispositifs hétérogènes reste encore une tâche complexe.

Dans le chapitre précédent, nous avons vu une première réponse à cette situation qui remplit les deux objectifs principaux : la simplification protocolaire et l'unification des technologies sous un plan de contrôle unique. L'architecture MPLS reprend quelques idées innovatrices et permet une première simplification protocolaire, assimilable à une intégration des architectures de niveaux 2 et 3 (par exemple IP et ATM), et fournit un plan de contrôle unique.

Néanmoins, cette architecture basée sur le concept de commutation d'étiquettes est limitée par l'hypothèse que l'on peut englober sous le terme de *commutation de paquets* : l'étiquette est codée explicitement quelque part dans le paquet, soit en utilisant un codage normalisé connu sous le terme



(a) (b)

Fig. 3.1: Evolution Historique et Simplification Protocolaire

« shim header », soit en utilisant les possibilités offertes par une technologie de transport particulière.

A cet égard, l'IETF est allé plus loin encore avec l'abstraction associée au paradigme de la commutation d'étiquettes. Les premiers efforts apparaissent avec le terme *Multi Protocol Lambda Switching* MPλS, concernant l'intégration de l'architecture MPLS et les réseaux optiques. L'idée est simple : *la longueur d'onde avec laquelle on transmet un bloc d'information détermine implicitement une étiquette.* Commuter une étiquette correspond à une commutation de longueur d'onde. Cette architecture est décrite dans la section suivante.

3.2. Une première approche, le MPλS

L'architecture connue sous le nom de MPλS fut motivée par l'intérêt de concevoir un plan de contrôle pour des brasseurs optiques (OCX) tout en tirant parti des travaux réalisés dans le contexte de l'architecture MPLS. Le plan de contrôle déployé dans les réseaux de transport optiques est fortement basé sur des mécanismes de gestion réseau qui présentent quelques inconvénients :

- Temps de configuration important.
- Convergence lente en cas de panne.
- Mise en place (approvisionnement) difficile.
- Nécessité d'interventions manuelles.
- Systèmes de gestion différents pour des équipements provenant de fournisseurs différents.

De nombreuses similitudes entre les deux architectures ont rapidement été identifiées. Citons entre autres :

- Du point de vue conceptuel, un LSP et un canal optique sont une route unidirectionnelle.
- Un LSR et un OCX font la séparation entre la plan de contrôle et le plan de données (plan usager)
- Le plan de données utilise l'étiquette entrante pour acheminer un paquet étiqueté entre un port d'entrée et un port de sortie. Un OCX utilise sa matrice de commutation pour connecter un «path Och» entre un port d'entrée et un port de sortie.

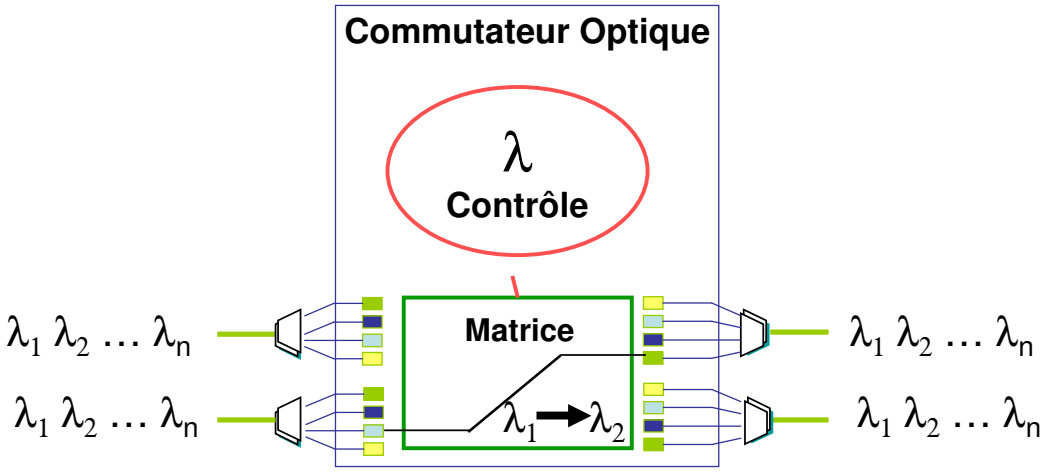


Fig. 3.2: Une première approche, le MPλS

Les équipements optiques sont capables d'établir des circuits de plus en plus rapidement. Certes, ces équipements ne sont pas des routeurs au sens où ils ne peuvent pas examiner les paquets entrants et choisir de manière dynamique le routeur suivant, mais l'intégration avec MPLS devient possible à partir du moment où l'on envisage la possibilité d'utiliser l'annonce d'une route vers une destination (préfixe donné) se propage à travers le réseau pour y établir un chemin optique *(ang. light path)*. L'ensemble des paquets destinés à un tel préfixe seront acheminés sur le chemin optique ainsi établi, sans la contrainte d'avoir à examiner l'information de niveau 3 à chaque nœud, procédure nécessitant une conversion O/E et E/O.

Les principaux atouts de MPλS sont les suivants :

- Un cadre pour la gestion de la bande passante et la création des canaux optiques, qui constitue un premier pas vers la normalisation des interfaces permettant d'offrir des services évolués de bande passante à la demande.
- L'intégration de l'Ingénierie de Trafic MPLS et du savoir-faire du routage IP.
- *La réutilisation des modules logiciels existants.*
- La coordination entre dispositifs à commutation de paquets et dispositifs optiques.
- La simplification de la gestion du réseau en fournissant une sémantique uniforme pour la gestion et le contrôle des réseaux aussi bien dans le domaine de la commutation de paquets que dans le domaine optique.

3.3. Vers la généralisation : GMPLS

La technologie MPλS a continué à évoluer vers ce que l'on nomme MPLS généralisé (GMPLS), architecture dans laquelle MPLS et MPλS apparaissent comme deux déclinaisons possibles d'une architecture plus générale. L'objectif de GMPLS est multiple : tout d'abord, identifier les opérations et

tâches concernant la commutation et l'acheminement d'informations susceptibles d'être vues comme des cas particuliers d'une formalisation associée à l'idée de commutation d'étiquettes. Deuxièmement, identifier les éléments nécessaires pour la prise en charge de telles opérations (extensions des protocoles de signalisation, prise en charge de LSPs bidirectionnels, etc.). Enfin, définir le contexte permettant le développement d'un plan de contrôle unique, où tous les équipements du domaine GMPLS (hétérogènes au niveau du plan usager, mais homogènes au niveau du plan de contrôle) sont capables d'interpréter les messages des protocoles de contrôle associés. Ainsi, la commutation d'étiquettes classique correspond au cas particulier où l'étiquette apparaît explicitement dans le bloc de données, et où elle est interprétée par l'équipement, qui est alors dit «Packet Switch Capable» ou PSC. D'autres cas possibles sont la commutation de longueur d'onde, définissant des équipements dits «Lambda Switch Capable» ou LSC, la commutation d'intervalles de temps dans une trame SDH, ou les équipements sont «Time Switch Capable» ou TSC et finalement la commutation spatiale de fibre optique avec des équipements dits «Fiber Switch Capable» ou FSC.

L'architecture GMPLS est née avec pour mission l'intégration des différents paradigmes de commutation présents dans un réseau à haut débit et l'unification du plan de contrôle permettant une simplification protocolaire. Bien évidemment, ceci impose un certain nombre de changements au niveau de l'opération du réseau.

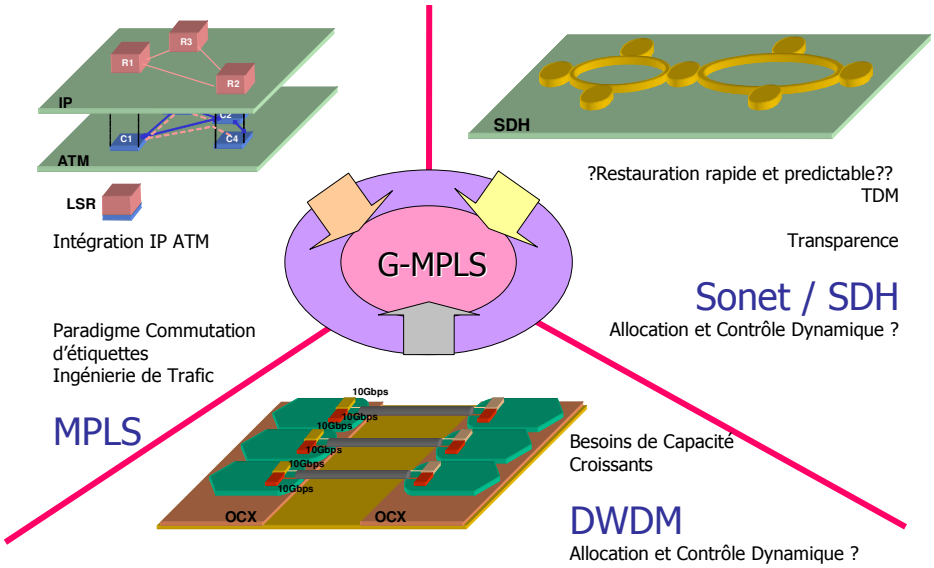


Fig. 3.3: Convergence vers GMPLS

3.4. Objectifs de GMPLS

D'une façon synthétique GMPLS permet :

- La réduction des coûts opérationnels et de gestion associés aux réseaux hétérogènes.
- Une meilleure intégration des domaines réseaux à commutation de paquets et des réseaux «tout optique».

Au moyen de :

– **L'intégration des différents paradigmes de commutation**

  La commutation d'étiquettes classique (sous forme d'objets binaires codés et ajoutés au bloc de données), la commutation de longueurs d'onde (objectif de MPλS) ou la commutation spatiale des intervalles de temps des réseaux synchrones deviennent des cas particulièrs d'une commutation abstraite.

– **L'unification du plan de contrôle**

  La définition d'un plan de contrôle et de gestion unifié.

– **L'ajout de nouveaux éléments**

  GMPLS requiert l'extension des protocoles de contrôle et de signalisation hérités de l'architecture MPLS, la généralisation de la notion d'étiquette, la prise en charge de LSPs bidirectionnels et l'utilisation de nouveaux protocoles (tels que LMP, ou Link Management Protocol).

– **Abstraction de la technologie**

  Les équipements réseaux sont hétérogènes du point de vue du plan de contrôle.

3.5. La hiérarchie GMPLS

La hiérarchie GMPLS généralise celle de MPLS : de façon similaire, la hiérarchie GMPLS est développée autour de la notion de LSP imbriqué, dans laquelle plusieurs LSPs sont agrégés dans un LSP trunk. Dans l'architecture MPLS, la hiérarchie était définie par rapport à une pile d'étiquettes qui était explicitement codée dans le paquet lui-même. La hiérarchie GMPLS est définie par rapport aux différentes technologies et capacités de commutation, qui vont définir implicitement comment les LSPs peuvent être imbriqués, de la façon suivante :

1. Un LSP commence et finit dans des équipements ayant une fonctionnalité identique.
2. Des LSPs ayant l'origine à un équipement capable de commuter des paquets (PSC) peuvent être imbriqués dans un LSP de type «TDM». Plusieurs LSPs de type «TDM» peuvent être à leur tour imbriqués dans un LSP de type «LSC-LSP» et ceux-ci, imbriqués dans un LSP de type «FSC-LSP». Ceci est possible car les extensions aux protocoles de routage vont considérer des LSPs comme des liaisons au niveau du routage, mais en respectant la hiérarchie définie.

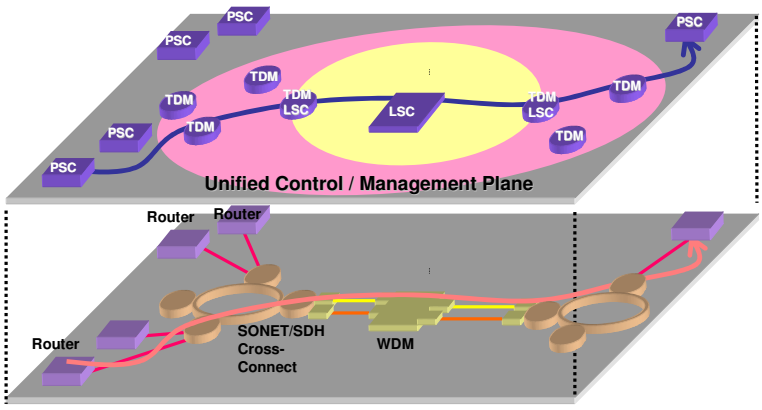
Imbrication de LSPs (ang. LSP nesting)

Une deuxième notion très liée à la notion de hiérarchie GMPLS est celle de l'imbrication de LSPs. Cette imbrication a lieu de façon naturelle dans la hiérarchie GMPLS, et elle est caractérisée par les propriétés suivantes :

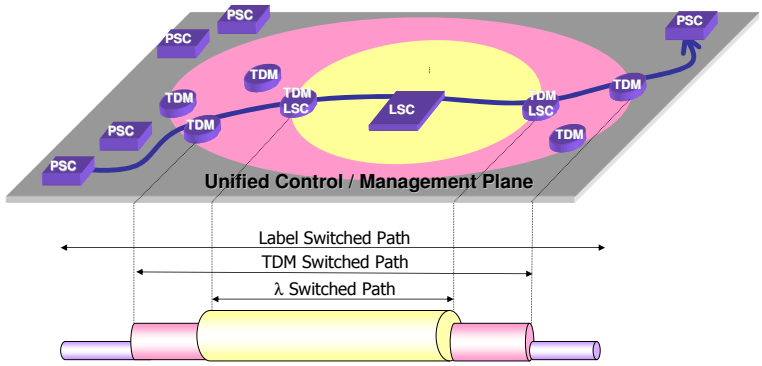
– Les canaux optiques ont des bandes passantes discrètes : OC-48, OC-192, OC-768
– A l'aide des LSP imbriqués, plusieurs LSPs sont concentrés dans un seul LSP optique, en tirant parti de l'agrégation et de la hiérarchie.

– Cette possibilité est une réponse au problème du gaspillage de la bande passante.

La figure 3.4 illustre la notion de hiérarchie GMPLS.



(a)



(b)

Fig. 3.4: La hiérarchie GMPLS

3.6. Extensions nécessaires

MPLS manque de certaines fonctionnalités nécessaires pour les objectifs cités. Les tunnels MPLS sont unidirectionnels, limitant l'application à des réseaux optiques. L'idée d'étiquette doit être généralisée, afin de prendre en charge les différents paradigmes de commutation existants. Les extensions nécessaires peuvent se résumer dans les catégories suivantes : l'extension des protocoles de routage et des fonctionnalités de gestion, l'extensions des protocoles de signalisation et l'introduction de nouveaux protocoles et de nouvelles fonctionnalités. Nous présentons dans la suite les éléments les plus

importants. Une analyse détaillée des extensions nécessaires en cours de normalisation est donnée dans [38] et [37].

Le déploiement de GMPLS suscite un certain nombre d'enjeux :

1. L'espace d'étiquettes MPLS est relativement grand. Par contre, le nombre de longueurs d'onde disponibles ou des intervalles de temps dans les réseaux SDH reste limité.
2. L'allocation de la bande passante à un LSP dans un contexte optique ou dans un contexte SDH doit être réalisée en utilisant un ensemble de valeurs finies.
3. Le nombre de liens gérés par deux LSRs adjacents peut être de l'ordre de plusieurs centaines (par exemple, on peut concevoir deux LSRs ayant plusieurs fibres en parallèle ou chaque fibre gère de centaines de longueurs d'onde). L'ordre de grandeur du nombre de liens à gérer par ces LSRs peut être beaucoup plus grand que dans un contexte MPLS classique.
4. Il n'est pas possible d'attribuer à chaque lien (localement) une adresse IP, non seulement parce que celles-ci sont limitées, mais aussi parce que une telle attribution est coûteuse en termes de gestion.

Les extensions que nous présentons dans la suite ont été conçues pour faire face à ces enjeux.

3.6.1. Extensions des protocoles de routage et des fonctions de gestion

Les protocoles de routage interne (IGP) tels que OSPF ou IS/IS sont en cours de normalisation en ce qui concerne leur extension pour la prise en charge de l'ingénierie de trafic [75]. Dans la section précédente, nous avons vu l'intérêt de la hiérarchie MPLS. La prise en charge d'une telle hiérarchie est réalisée en annonçant des LSPs actifs comme des liens dans le protocole de routage IGP, respectant l'ordre hiérarchique défini. L'utilisation de ces protocoles dans un contexte GMPLS nécessite la définition d'un certain nombre de TLVs (en cours de normalisation), comme le *Link Type TLV* pour identifier le type de lien et ses capacités de commutation ou le *Shared Risk Link Group (SRLG) TLV* destiné aux mécanismes de protection et à la tolérance aux pannes. En résumé, les extensions des fonctionnalités de routage sont : (a) la définition de nouveaux TLV et de nouveaux attributs des liens comme le *type,* (b) l'assimilation des LSPs à des liens logiques dont les informations topologiques et attributs sont redistribués par les protocoles de routage, en respectant la hiérarchie définie, (c) la notion du regroupement de liens et (d) la gestion des liens non numérotés.

Adjacence d'acheminement (ang. Forwarding adjacency)

Les LSPs sont annoncés par le protocole de routage IGP et donc pris en compte par les différentes heuristiques de calcul de routes. Notons que ces heuristiques (comme CSPF, présentée au chapitre suivant) devront aussi être adaptées au contexte GMPLS.

Regroupement de liens (ang. Link bundling).

– Des nœuds peuvent être connectés par des centaines de liens en parallèle (citons par exemple, les longueurs d'onde).
– Le regroupement de liens permet que plusieurs liens soient annoncés comme un seul lien par le protocole IGP.

– Cela permet un meilleur passage à l'échelle en réduisant la quantité d'informations que OSPF ou IS/IS doit prendre en charge.

Liens non numérotés (ang. unnumbered links).

L'établissement d'un LSPs nécessite la connaissance des LSRs et des liens traversés. L'attribution d'une adresse IP à chaque interface connecté à un lien peut ne pas être envisageable quand les LSRs sont connectés par un nombre important de liens (citons par exemple, N longueurs d'onde possibles sur une même fibre optique). Une solution possible à ce problème consiste à attribuer à chaque nœud ou LSR un identificateur unique (on utilise typiquement une adresse IP de loopback comme identificateur du nœud) et ensuite identifier localement chacun de ses liens. Pour ceci, les protocoles de routage et les protocoles de signalisation doivent être étendus pour la prise en charge de ce type de liens que l'on nomme des «liens non numérotés» . Ainsi, un lien est globalement identifié à l'aide du couple *(LSR Id, Link Number).* Les avantages d'une telle approche sont (a) la réduction des coûts de gestion associés à l'adressage IP et (b) la duplication d'adresses.

3.6.2. Extensions des protocoles de signalisation

Les protocoles de signalisation tels que RSVP-TE ou CR-LDP doivent aussi être étendus pour la prise en charge de GMPLS. D'une façon succincte, les principaux éléments à ajouter actuellement en cours de normalisation sont les suivants :

– Prise en charge de l'établissement de LSPs hiérarchiques.
– Normalisation de la notion d'Etiquette Généralisée (intégrant les différents paradigmes de commutation).
– Généralisation du message de *Label Request.*
– Prise en charge de LSPs bidirectionnels, pour une meilleure intégration avec les réseaux optiques.
– Notion d'*étiquette suggérée.*
– Utilisation des *ensembles et espaces d'étiquettes.*
– Généralisation des mécanismes de protection et de restauration.
– Généralisation des messages de notification.

3.6.3. Nouveaux Protocoles et fonctionnalités

Voie de Contrôle (ang. Control Channel)

L'architecture GMPLS normalise la notion de *Voie de Contrôle,* qui présente les propriétés suivantes :

– Destinée à l'échange d'informations entre LSRs adjacents, notamment pour la signalisation, le routage et la gestion.
– Peut utiliser une voie physique différente de celle des données (ang. In-Band, Out-Band).

Protocole de Gestion de Liens ou Link Management Protocol (LMP)

Parmi les nouvelles fonctionnalités nécessaires dans l'architecture GMPLS, le protocole *Link Management Protocol* ou LMP est particulièrement important. La normalisation de LMP a été motivée par la constatation que dans le contexte GMPLS, deux LSR adjacents peuvent être connectés par un

nombre relativement important de liens en parallèle. Ceci pose un certain nombre de problèmes de passage à l'échelle, résolus dans une certaine mesure grâce à la notion de Link Bundling. Le protocole LMP, exécuté entre deux nœuds adjacents assure quatre fonctionnalités principales :

- Permet l'établissement et la gestion de la voie de contrôle. Sur cette voie, une instance d'un protocole simple de type *Hello Keep-Alive* est exécutée. Remarquons que la voie de contrôle peut être séparée des liens composants, par exemple, en utilisant une connexion de type Ethernet dédiée.
- Réalise la vérification et la gestion de la connectivité des canaux de transmission de données d'usager *(bearer channels)*
- Corrélation des propriétés des liens.
- Isolation de pannes.

3.7. Modèles de Déploiement

Plusieurs modèles pour l'architecture GMPLS sont susceptibles d'être déployés : le modèle superposé (ou Overlay) et le modèle intégré (ou Peer).

3.7.1. Modèle Superposé ou Overlay

Dans le modèle Overlay (proposé par l'organisme Optical Interworking Forum, OIF) le réseau client est, au niveau du routage, indépendant du réseau optique. Le réseau de transport est caché, et deux plans de contrôle indépendants coexistent avec une interaction minimale. Les fonctions d'interfonctionnement sont réalisées en définissant des interfaces, et pour cela deux interfaces sont normalisées, l'interface UNI *(User to Network Interface)* et l'interface NNI *(Network to Network Interface).* L'interface UNI, (par exemple, entre le routeur de sortie du domaine et le dispositif optique) permet aux équipements clients de réaliser des requêtes d'établissement de tunnels optiques, dont le réseau client bénéficie de façon transparente. La découverte des voisins, la distribution des informations topologiques (routage) et les protocoles de signalisation sont indépendants.

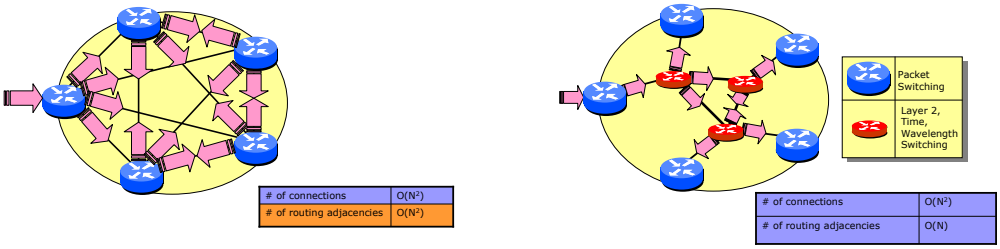
Le modèle Overlay est illustré dans la figure 3.5(a).

3.7.2. Modèle Intégré ou Peer

Le modèle Peer est proposé par l'IETF. Dans ce modèle, tous les dispositifs sont censés exécuter une ou plusieurs instances d'un protocole de routage (par exemple, O-OSPF), qu'ils soient optiques ou à commutation de paquets. Les routeurs MPLS agissent comme des peers (partenaires) des commutateurs optiques. Les informations de routage sont échangées entre eux, et donc tous les équipements ont une connaissance plus ou moins détaillée de la topologie du réseau.

Bien évidemment, le modèle peer implique l'extension et la généralisation des protocoles existants. Il paraît raisonnable que dans un contexte réseau où le protocole IP est le protocole dominant, les protocoles de routage développés à ce propos soient des versions étendues de protocoles IGP classiques tels que OSPF ou IS/IS.

Le modèle Peer est illustré dans la figure 3.5(b). L'avantage du modèle peer par rapport au modèle superposé est la possibilité d'établir $O(N^2)$ chemins de données en gérant seulement $O(N)$ adjacences de routage.



(a) (b)

Fig. 3.5: Modèles de déploiement de GMPLS

3.7.3. Remarques

Remarquons que le modèle Peer est un sur-ensemble du modèle Overlay. L'ensemble des fonctionnalités nécessaires pour la prise en charge du modèle Overlay est inclus dans celui pour la prise en charge du modèle Peer. Concrètement, le modèle Overlay peut être implémenté en désactivant les fonctionnalités d'échange de topologie tout en gardant les fonctionnalités de signalisation. Il existe d'autre part des modèles hybrides.

3.8. Conclusions

L'architecture GMPLS est très prometteuse, et permet l'intégration sous un seul plan de contrôle de technologies hétérogènes. A ce propos, l'architecture MPLS est devenue une déclinaison possible d'une architecture plus générique, GMPLS, dont les objectifs sont bien définis et justifiés mais, étant en cours de normalisation, le nombre d'implémentations est très réduit, et de nombreuses questions autour de GMPLS restent ouvertes.

4. Eléments d'Ingénierie de Trafic

4.1. Introduction et Motivation

Dans les deux chapitres précédents nous avons présenté l'architecture MPLS et comment celle-ci fournit de nouveaux mécanismes pour l'ingénierie de trafic. Dans les sections suivantes, nous allons détailler les approches existantes pour mettre en place l'ingénierie de trafic, ainsi que leurs relations avec les plans de contrôle et de gestion. Sauf mention explicite, nous nous plaçons dans le contexte IP sur MPLS.

La structure du chapitre est la suivante : la notion d'Ingénierie de Trafic est détaillée dans la section 4.2, et la section 4.3 énumère les différents composants d'ingénierie de trafic du plan de contrôle de MPLS, qui sont détaillés ensuite : la découverte des ressources (section 4.4), la diffusion de l'état du réseau (section 4.5), le calcul et sélection de routes (section 4.6), l'assignation du trafic (section 4.7) et les mécanismes de restauration (section 4.8). Un exemple de serveur pour l'ingénierie de trafic MPLS est donné dans la section 4.9.

La section 4.10 présente la notion de partage de charge, les mécanismes implémentés par les protocoles de routage IP classiques pour le partage de charge et leurs limitations. Ensuite, nous formulons le problème du partage de charge dans les réseaux MPLS, et nous donnons un aperçu de l'état de l'art et des travaux existants ainsi que de certains aspects architecturaux du partage de charge et de leur état de normalisation. Finalement la section 4.11 conclut le chapitre.

4.2. Ingénierie de Trafic

D'après la définition donnée dans [88], l'ingénierie de trafic a pour but l'optimisation, du point de vue des performances, des réseaux opérationnels. D'une façon générale, l'ingénierie de trafic englobe l'application de principes technologiques et scientifiques à la mesure, la modélisation, la caractérisation et le contrôle du trafic et l'application de ces principes et du «savoir-faire» correspondant pour obtenir des objectifs de performance concrets.

Historiquement, l'ingénierie de trafic dans l'Internet était limitée à la gestion isolée des files d'attente des routeurs et des éléments réseaux et à l'approvisionnement manuel de routes alternatives à celles imposées par le protocole de routage, en utilisant par exemple des routes statiques. Ceci est principalement expliqué par le fait que les protocoles de routage étaient (et dans un certain sens restent) limités en ce qui concerne la prise en charge des extensions pour l'ingénierie de trafic. Une conséquence d'une telle limitation est que pour un couple origine / destination donné, l'utilisation de l'acheminement classique IP restreint le trafic à la route dite *de plus court chemin,* ou route par défaut.

La notion de FEC, introduite dans l'architecture MPLS, fournit une meilleure granularité pour la répartition du trafic que l'approche IP classique qui utilise uniquement l'adresse destination.

Les protocoles de signalisation associés [12] [95] viennent compléter les blocs fonctionnels de cette architecture en permettant l'établissement de routes autres que celles définies par le protocole de routage. Enfin, la possibilité d'une réservation explicite de ressources pour un agrégat de trafic ou *traffic trunk* est une façon de lui garantir un certain degré de qualité de service.

Le RFC 2702, section 3.2 identifie trois problèmes fondamentaux de l'ingénierie de trafic dans les réseaux MPLS : partitionner le trafic en classes d'équivalence pour l'acheminement (FECs), associer ces FECs à des LSPs et, finalement, faire correspondre ces LSPs à la topologie physique du réseau (layout de LSPs).

La figure 4.1 illustre le problème du *layout* : Faire correspondre les LSPs à la topologie physique mise en place.

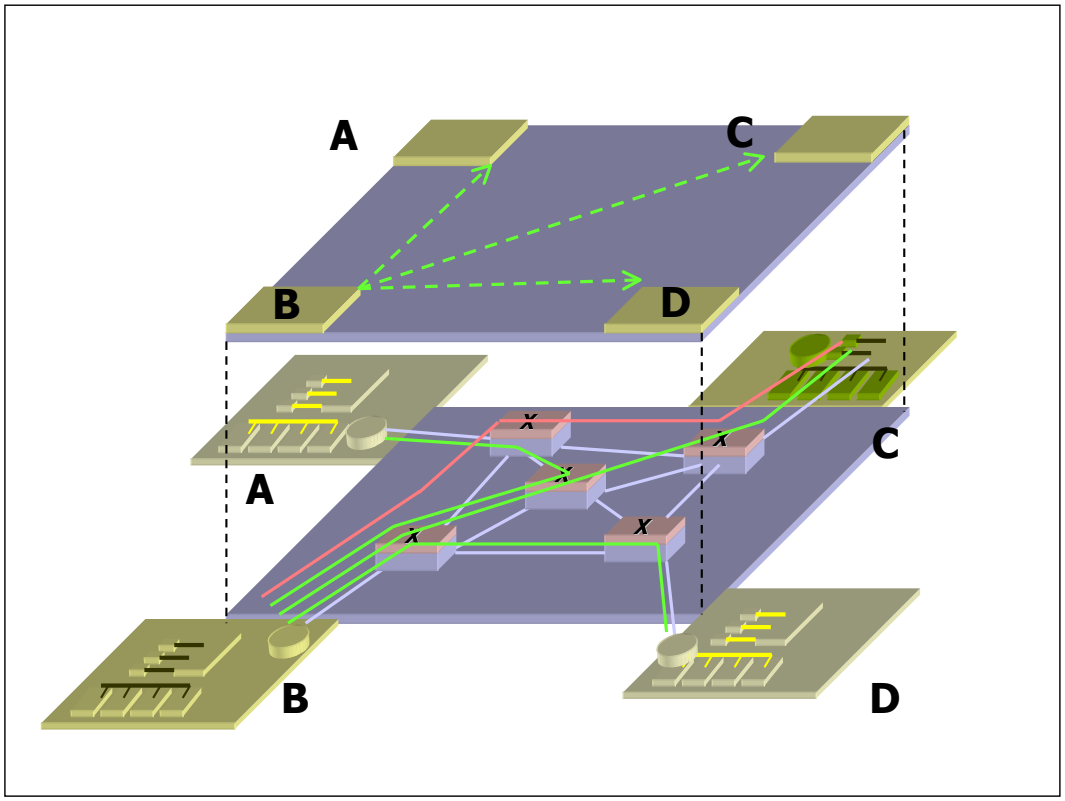


Fig. 4.1: LSP layout : correspondance entre la matrice de trafic et la topologie physique du réseau.

4.3. Composants d'Ingénierie de Trafic du plan de contrôle de MPLS

Découverte de Ressources La découverte et la gestion des ressources disponibles par chaque équipement réseau est l'étape préalable à la diffusion aux entités concernées de l'état du réseau.

Cette découverte est réalisée automatiquement (p.e. pour des ressources configurées manuellement) ou suite à la réception de messages des protocoles de contrôle.

Diffusion de l'état du réseau La diffusion de l'état du réseau est faite à l'aide des protocoles de routage existants, éventuellement modifiés et étendus. Les informations concernées sont la topologie et la disponibilité de ressources comme la bande passante nominale ou résiduelle. Ces messages sont appelés de façon générique (par abus du langage suite à leur définition dans un protocole particulier) des messages de notification d'état des liens *(ang. Link State Advertisements ou LSA)* .

Sélection de Route Le terme *calcul et sélection de route* englobe les procédures, les algorithmes et les heuristiques destinés au calcul d'une structure de données contenant la liste de routeurs entre deux éléments du réseau. Cette liste vérifie un certain nombre de contraintes administratives et de qualité de service. D'une façon générale, la complexité des algorithmes destinés à un tel calcul devient intraitable lorsque la taille du réseau et le nombre de contraintes augmente. Une heuristique pour le calcul de routes largement utilisée est l'heuristique du «Plus Court Chemin Contraint» *(ang. Constrained Shortest Path First - CSPF)* [88].

Gestion des Routes La gestion de routes concerne les mécanismes, algorithmes, procédures et protocoles permettant l'établissement, la gestion et le relâchement de routes (LSPs) dans un domaine MPLS, et notamment la distribution d'étiquettes. Comme évoqué précedemment, deux protocoles sont aujourd'hui normalisés et utilisés à cette fin : CR-LDP [12] et RSVP-TE [95].

Assignation du trafic L'assignation de trafic est composée de deux fonctionnalités : la fonction de partition et la fonction d'attribution. Ceci est décrit dans la section 4.7.

Mécanismes de Protection et de Restauration Les mécanismes de protection et de restauration assurés par les protocoles de routage dynamiques existants peuvent s'avérer insuffisants pour certaines applications. L'architecture MPLS apporte des mécanismes nouveaux visant une restauration en des temps de l'ordre de quelques millisecondes, correspondant aux critères et exigences actuels. Remarquons qu'il ne s'agit pas seulement de proposer de nouveaux mécanismes pour la protection et la restauration mais aussi de les synchroniser avec les autres mécanismes fournis par les différentes couches protocolaires.

4.4. Découverte de Ressources

Certaines ressources dites locales aux routeurs sont découvertes automatiquement. Citons par exemple les différentes interfaces réseaux et leurs adresses IP. La découverte des voisins topologiques est réalisée grâce à l'échange de messages de découverte, designés, par abus de langage sous le nom de *Messages de Hello,* envoyés à des adresses de diffusion normalisées et spécifiques aux protocoles de routage.

4.5. Diffusion de l'état du réseau

Un certain nombre de procédures nécessitent une connaissance plus ou moins détaillée de l'état du réseau (l'exemple classique est la sélection d'une route explicite suite une requête d'établissement d'un chemin). Les métriques statiques s'avèrent parfois insuffisantes lorsque le réseau doit fournir des services avec un certain niveau de qualité. Les mécanismes et protocoles de routage dynamiques de type «à état de liens» sensibles à la qualité de service[1] tels que OSPF-TE, Q-OSPF ou ISIS-TE sont une façon (non exclusive) de garantir la diffusion des informations susceptibles d'être prises en compte dans les différentes heuristiques (cf. section 4.6.3.1). La diffusion de l'état du réseau est réalisée grâce à l'échange de messages (par exemple, le protocole OSPF [74] définit les messages LSA (*Link State Advertisements*) permettant la synchronisation, entre les LSR appartenant à un domaine, de leurs bases de données topologiques.

Caractérisation des Liens

Chaque lien du réseau est caractérisé par ses métriques d'état. Certains travaux font l'hypothèse que chaque nœud du réseau a une connaissance parfaite de l'état dynamique du réseau, grâce à un protocole de routage sensible à la QoS. Les effets dus à la fréquence de rafraîchissement des mises à jour, les différents délais de transmission et de propagation des messages des protocoles de routage et d'une manière générale, les effets de la synchronisation ne sont pas toujours pris en compte. Il est souvent admis que si les changements des modèles de trafic sont lents par rapport au délai maximal d'un transfert de bout en bout, ces approches restent valides.

4.5.1. Le routage hiérarchique

Le routage dit hiérarchique est une réponse à l'explosion des informations de la base de données topologique du routage et permet d'agréger de façon élégante ces informations lorsque la taille du réseau est importante. Considérons par exemple le routage hiérarchique des réseaux ATM (cf. [52]) : afin de simplifier les bases de données topologiques et les tables de routage, chaque nœud gère des informations topologiques détaillées concernant le plus bas niveau hiérarchique auquel il appartient, ainsi que des informations agrégées concernant le reste du réseau. Ceci nécessite la définition de groupes de nœuds, partageant un attribut commun (géographique, administratif, etc.)

Un modèle souvent utilisé pour l'évaluation des performances du routage hiérarchique [5], [60], [6] est le suivant : le réseau physique est représenté par un graphe orienté, représentant le niveau ou couche 0. Les nœuds du niveau 0 sont groupés en grappes *(ang. cluster)*. Ces clusters deviennent les nœuds logiques du niveau 1, etc. Autrement dit, le système est la superposition d'un ou plusieurs niveaux hiérarchiques, chaque niveau étant constitué de nœuds logiques interconnectés par des liens logiques. Le niveau 0 correspond à la topologie physique. Chaque nœud a une vision topologique différente, selon le groupe ou cluster auquel il appartient. Dans les réseaux ATM [2]utilisant PNNI [82] comme protocole de routage, les clusters sont nommés *Peer Groups* .

Le nombre optimal de niveaux hiérarchiques est une question difficile, qui dépend fortement du

[1]en anglais, QoS-aware link state routing protocols

[2]Comme dans les réseaux MPLS, on parle de réseau ATM comme un abus de langage pour référencer les réseaux qui utilisent ATM *(Asynchronous Transfer Mode)* comme technologie de transfert.

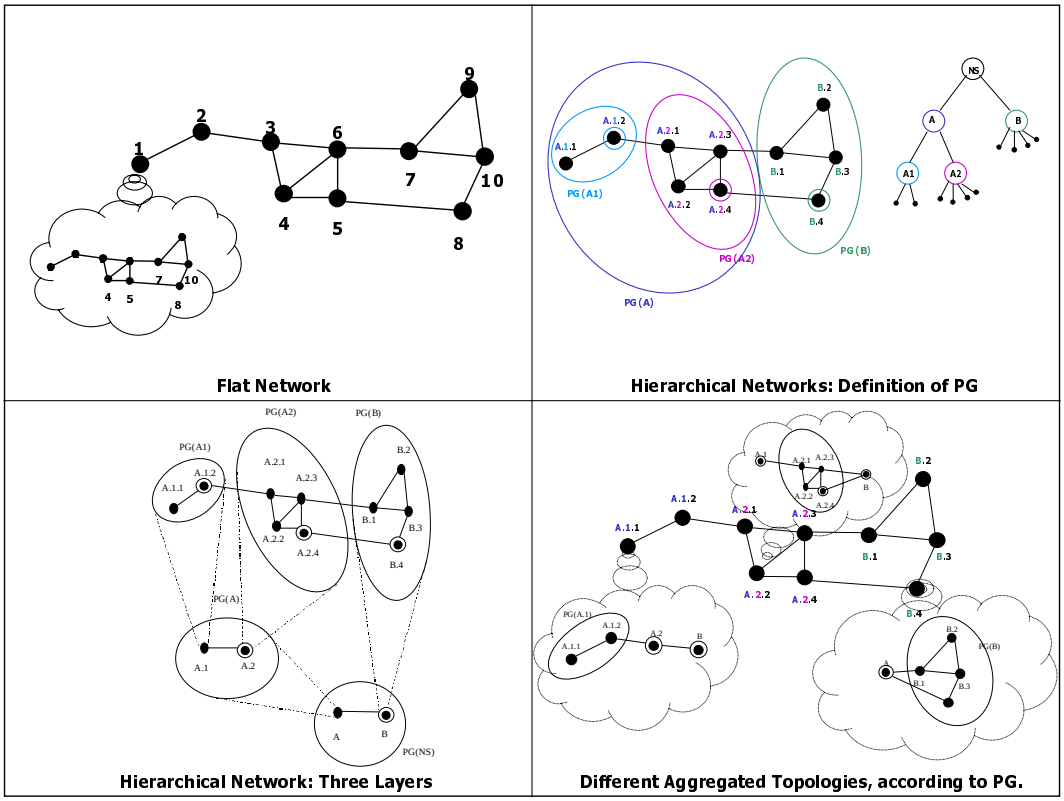


Fig. 4.2: Modélisation du réseaux comme un graphe.

critère considéré. Avec un nombre limité de niveaux, les effets de la hiérarchisation du routage ne sont pas notables, et avec un nombre trop élévé, la complexité de gestion devient trop importante pour un bénéfice négligeable [52]. Citons par exemple les 3 niveaux de la hiérarchie de l'Internet actuel (en Systèmes Autonomes et en aires de routage OSPF). La notion de routage hiérarchique, dans le contexte des réseaux ATM / PNNI est illustrée sur la figure 4.2. On peut noter comment la quantité d'informations de routage gérés par chaque nœud diminuent.

Des études concernant la performance des différentes méthodes d'agrégation d'information sont présentées dans [6], [71], [72]. Des techniques de passage à l'échelle sont décrites dans [43], [44], [45]. Finalement, une approche pour le calcul distribué de routes dans les réseaux ATM est présentée dans [31].

4.6. Calcul et Sélection de Route

4.6.1. Définition

Le terme «calcul des routes» *(Path computation),* et son extension le «calcul de routes multiples» *(Multipath computation)* définissent l'ensemble des procédures, algorithmes et heuristiques ayant

comme objectif l'identification et la classification des chemins constituant un «groupe de LSPs» ou *Path Group* entre deux nœuds, normalement, un routeur d'entrée I-LSR et un routeur de sortie E-LSR, dans un domaine MPLS. La problématique relative à ce calcul est un sujet de recherche actuel non seulement dans le contexte des réseaux à commutation de paquets mais aussi dans le contexte du contrôle automatique et dans d'autres domaines. Les outils mathématiques utilisés font souvent appel à la théorie des graphes, avec parfois des approches probabilistes. Il est connu que même avec des hypothèses simples (par rapport aux ambitions humaines), les problèmes associés sont complexes (NP-complets avec deux métriques non concaves). Plusieurs heuristiques sont alors proposées, favorisant parfois la vitesse de l'algorithme ou la prise en compte de contraintes de puissance de calcul ou de taille mémoire en détriment de la sous-optimalité de la solution.

4.6.2. Les questions ouvertes

Le calcul des routes tente de répondre à des questions concernant le nombre de routes à établir, les paramètres importants (bande passante, gigue, délai,...), les entités responsables du calcul et comment celui-ci est réalisé (avec une architecture centralisée, distribuée, mixte...), ainsi que la validité temporelle du résultat : configuration manuelle indéfinie, établissement dynamique sur demande, etc.

4.6.3. Algorithmes et heuristiques

Une synthèse exhaustive de la complexité des algorithmes les plus utilisés pour le calcul des routes avec des contraintes de qualité de service est donnée dans [104] et [14]. [3], [98], [5], [67] analysent des algorithmes destinés au calcul efficace de chemins «sur demande» avec ou sans contraintes de bande passante, et présentent l'état de l'art concernant certains algorithmes comme *Widest Shortest Path, Shortest Widest Path* et des variantes existantes sur les métriques utilisées. Finalement, Oueslati [80] développe ses travaux autour de la qualité de service et routage des flots élastiques dans un réseau multiservice. L'auteur compare les performances de plusieurs algorithmes de routage adaptatif et propose celui de *Maximum Utility.*

4.6.3.1. L'heuristique MPLS / CSPF

Le but de cette section est de présenter l'heuristique CSPF *(ang. Constrained Shortest Path First),* conçue pour le calcul de routes respectant des contraintes, qui a été reprise par un grand nombre de constructeurs du fait qu'elle présente un bon compromis entre efficacité et optimalité.

Définition 4.6.1 (Graphe aminci sur les liens). Etant donné un graphe G=(V,E) et une application faisant correspondre à chaque lien $l \in E$ une valeur dans $\{0, 1\}$ on appelle un graphe aminci sur les liens *(ang. link constrained graph)* le graphe obtenu en supprimant de G *(ang. prune)* les liens dont l'image vaut zéro. Un exemple classique est l'amincissement d'un graphe dans lequel on retire les liens pour lesquels la bande passante résiduelle est inférieure à une certaine valeur.

Définition 4.6.2 (Graphe aminci sur les noeuds). Etant donné un graphe G=(V,E) et une application faisant correspondre à chaque nœud $n \in V$ une valeur dans $\{0, 1\}$ on appelle un graphe aminci sur les nœuds *(ang. node constrained graph)* le graphe obtenu en supprimant de G *(ang.*

prune) les nœuds dont l'image vaut zéro, ainsi que les liens (entrants / sortants) correspondants *(ang. incoming and outgoing links).* Un exemple est l'amincissement qui à lieu au niveau des nœuds logiques représentant des Systèmes Autonomes, suite à l'application de contraintes politiques et administratives.

L'heuristique MPLS / CSPF consiste à appliquer itérativement des contraintes de caractère politique, économique, technologique, administratif, etc., où chaque contrainte induit un amincissement du graphe modélisant le réseau. L'algorithme de Dijkstra calcule, sur le sous graphe résultat, le plus court chemin par rapport à une métrique donnée. Remarquons que les sous graphes résultants à chaque itération peuvent ne pas être connexes.

4.7. Assignation du Trafic

Cet aspect de l'ingénierie de trafic est divisé en deux fonctions : la fonction de partition et la fonction d'attribution.

4.7.1. Fonction de Partage ou Partition (ang. Partition Function)

Le trafic à transporter dans le réseau (la matrice de trafic) est divisé en classes d'équivalence ou FECs. La notion de FEC fournit une flexibilité et une granularité notable.

4.7.2. Fonction d'Attribution (ang. Apportionment function)

La fonction d'attribution fait correspondre à chaque classe d'équivalence un ou plusieurs chemins. Le fait d'attribuer à une FEC un nombre arbitraire de LSPs nous permet de mettre en place des mécanismes de partage de charge, notion sur laquelle nous reviendrons dans la section 4.10.

4.8. Les mécanismes de restauration basés sur MPLS

Afin de fournir un service fiable, l'architecture MPLS a besoin d'un ensemble de procédures destinées à assurer la protection du trafic transporté sur les LSPs[3]. Ceci impose que les routeurs faisant parti du domaine MPLS fournissent des mécanismes *de détection, de notification et de restauration* des pannes. De plus, la signalisation associée doit aussi supporter ces extensions. La protection du trafic dans le contexte MPLS, connue sous le terme générique de «Restauration MPLS» *(ang. MPLS-based Recovery)* présente des avantages importants : D'une part, la possibilité d'améliorer la fiabilité du réseau en permettant une réaction aux pannes plus rapide qu'avec une architecture IP classique. D'autre part, les mécanismes de protection MPLS peuvent s'avérer utiles dans le cadre d'architectures IP sur canaux optiques WDM, en évitant une couche SDH.

Justification des mécanismes

– Le re-routage de niveau 3 (re-routage IP) peut s'avérer trop lent dans le cœur d'un réseau MPLS nécessitant une grande fiabilité et une grande disponibilité. Certes, les technologies comme DWDM ou

[3]Le groupe de travail de l'IETF a récemment commencé les travaux de normalisation autour des mécanismes de restauration basés sur MPLS. Les drafts sont récents, mais la terminologie utilisée semble stable.

SDH peuvent fournir des mécanismes de protection. Néanmoins, la synchronisation inter-couches correcte de ces mécanismes et la gestion d'une redondance contrôlée restent des questions ouvertes.
– La granularité avec laquelle les couches inférieures sont capables de protéger le trafic peut s'avérer trop grossière.
– La visibilité des mécanismes des couches inférieures au niveau des couches supérieures est réduite.
– L'interopérabilité des mécanismes de protection entre des équipements de fournisseurs différents est nécessaire afin d'adopter MPLS comme technologie de transport permettant une ingénierie de trafic.

4.8.1. Modèles de Restauration

Les protocoles de routage classiques sont relativement robustes, mais le temps qu'ils nécessitent pour rétablir le bon fonctionnement du réseau suite à une panne peut être important, de l'ordre de plusieurs secondes ou minutes, pouvant entraîner une interruption sérieuse du service pour les applications. Afin d'assurer un service fiable et tolérant aux pannes, les temps de restauration doivent être de l'ordre de la dizaine de millisecondes. Etant donné que les temps de restauration des protocoles de routage actuels ne peuvent être réduits du fait de limitations intrinsèques de ces protocoles, les mécanismes de restauration basés sur MPLS peuvent améliorer cette situation. Deux modèles sont aujourd'hui proposés et en cours de normalisation : la restauration par re-routage et la protection par commutation.

Ces deux modèles qui seront décrits dans la suite, ne sont pas mutuellement exclusifs. Par exemple, la protection par commutation peut être utilisée pour une restauration rapide garantissant la connectivité pendant que les mécanismes de re-routage déterminent une nouvelle configuration du réseau, avec un nouveau calcul des chemins si nécessaire.

Restauration par Re-Routage

La restauration par re-routage est définie comme l'établissement de nouveaux chemins (ou morceaux de chemin) sur demande afin d'assurer la restauration du trafic suite à l'occurrence d'une panne. Les nouveaux chemins peuvent utiliser des informations concernant la panne elle-même, concernant des politiques administratives et des politiques de routage, des configurations prédéfinies et des informations topologiques du réseau. Ainsi, suite à la détection de la panne, les chemins ou morceaux de chemin utilisés pour contourner la panne en détournant le trafic sont établis par signalisation. Les mécanismes de re-routage sont intrinsèquement plus lents que les mécanismes de protection par commutation définis dans la suite, du fait de la procédure à suivre suite à la détection d'une panne. Néanmoins, ils s'avèrent moins gourmands en ressources, car ils ne nécessitent pas de réservation de ressources avant l'occurrence de la panne ni la connaissance de celle-ci. Une fois que les protocoles de routage ont convergé, il peut être préférable de ré-optimiser le réseau en réalisant un re-routage d'après le nouvel état du réseau et d'après toute politique existante. *La restauration par re-routage utilise des chemins établis par signalisation sur demande avec réservation de ressources sur demande. Les chemins (ou morceaux de chemin) sont calculés dynamiquement.*

Restauration par Protection par Commutation

Les mécanismes de restauration MPLS dits de «Protection par Commutation» *(ang. Protection Switching)* pré-établissent des chemins (ou morceaux de chemin) de secours, selon des politiques de routage, selon les besoins de restauration du trafic à acheminer sur les chemins principaux et selon les politiques administratives. Les chemins de secours peuvent être ou non disjoints lien-à-lien ou nœud-à-nœud avec le chemin principal. Néanmoins, si les chemins de secours partagent avec les chemins principaux des sources potentielles de panne, la tolérance du réseau aux pannes en est réduite. Suite à la détection d'une panne, le trafic est acheminé sur le ou les chemins de secours et restauré. *La Protection par Commutation utilise des chemins de secours préétablis. Si une réservation de ressources est nécessaire, la protection par commutation utilise aussi des ressources réservées à l'avance.*

4.8.2. Notification des Pannes

Les mécanismes de protection par commutation nécessitent une notification des pannes rapide et fiable. Le noeud détectant une panne doit décider si la celle-ci est suffisamment grave pour justifier le déclenchement de la restauration. Le noeud doit notifier la détection de la panne au routeur upstream en envoyant une FIS *(Fault Indication Signal).* Ceci signifie que tout routeur capable de détecter une panne doit gérer une liste des routeurs amont. Cette notification va monter pas-à-pas jusqu'à arriver à un élément du réseau capable de prendre une décision. On appelle cet élément un PSL *(PSL ou Path Switch LSR).* Seul un PSL peut terminer une FIS.

Etant donnée que le FIS est un message de contrôle, il doit être envoyé avec une priorité haute afin d'arriver rapidement aux PSLs concernés.

4.8.3. Opérations de «Switch Back»

Les drafts actuels identifient deux modes d'opération suite à la détection d'une panne : réversible et non-réversible *(ang. revertive / non revertive).* Dans le mode réversible, on identifie un état de référence. Suite à une panne, le trafic est acheminé sur des liens ou chemins de secours. Lorsque cette panne est identifiée et réparée, le réseau revient *(switch back)* à l'état préféré. Par contre, dans le deuxième mode, le réseau ne revient pas au mode préféré.

4.9. RATES : Un exemple de serveur pour l'ingénierie de trafic MPLS

RATES est une architecture logicielle, ayant comme objectif la migration vers une architecture capable de gérer l'Ingénierie de Trafic de façon simple [87]. C'est une architecture proposée par Bell Labs, consistant en une base de données qui gère les politiques et les différentes contraintes, une interface basée sur un client navigateur Web permettant la définition et la manipulation de politiques et la gestion des ressources avec une approche client / serveur COPS. Cette approche permet l'échange des routes et des informations concernant les ressources disponibles entre le serveurs et les dispositifs de la périphérie. Le serveur RATES exécute une instanciation du protocole de routage OSPF lui permettant d'avoir une connaissance de l'état du réseau, et le calcul des routes

est basée sur un algorithme propriétaire dit de «Routage à Inférence Minimale», ayant pour objectif une utilisation optimale des ressources réseau.

4.10. Partage de charge, Première visite

4.10.1. Introduction et Motivation

Du fait de la nature «pas-à-pas» *(ang. «hop by hop»)*du routage IP, les protocoles de routage existants permettant le partage de charge (comme ECMP [89][90] ou OSPF-OMP) présentent certaines limitations nécessaires afin d'éviter les boucles de routage (par exemple, ECMP ne permet le partage de charge qu'entre des chemins de longueur égale) et leur fonction de partition de trafic dépend uniquement de métriques administratives ou de paramètres statiques comme la capacité nominale des liens de transmission.

La notion de *Classe d'Equivalence pour l'Acheminement (Forwarding Equivalence Class, FEC)* constitue un élément fondamental de l'architecture MPLS. Cette notion fournit une granularité et une souplesse importante pour partitionner le trafic selon un modèle de service choisi. Dans un contexte IP, la FEC la plus élémentaire correspond au «préfixe d'adresse destination», c'est-à-dire que deux paquets seront classés dans la même FEC et en conséquence traités de la même façon au niveau de l'acheminement s'ils ont le même préfixe d'adresse. Le plan de contrôle de MPLS a pour fonction l'établissement et le relâchement, à l'aide des différents protocoles de signalisation, des tunnels ou *Label Switched Paths (LSPs)* entre les points d'entrée (routeurs ingress) et les points de sortie (routeur egress). L'ingénieur réseau dispose ainsi de nouveaux mécanismes pour optimiser l'utilisation des ressources du réseau et offrir une certaine qualité de service en assignant des routes différentes à des classes de service différentes. En ce sens, le partage de charge peut être vu comme une extension de ces mécanismes consistant à acheminer une même FEC sur un ensemble de LSPs. Cet ensemble sera nommé le «groupe de LSPs» (ce terme sera définit de manière formelle ultérieurement).

Du fait de à la nature orientée connexion de l'architecture MPLS, plusieurs algorithmes de partage de charge peuvent être implémentés. La suppression de la limitation «entre chemins à coûts égaux» associée aux mécanismes de partage de charge classiques des protocoles de routage IP nous incite à chercher des éléments permettant d'évaluer d'une façon rigoureuse l'impact des propriétés stochastiques du trafic sur la performance des différentes façons de réaliser un partage de charge.

Bénéfices du partage de charge

On peut identifier plusieurs bénéfices importants au partage de charge : d'une part, le partage de charge réduit la congestion et les pertes, car par définition il permet la répartition du trafic. D'autre part, il offre un temps de réponse plus petit en cas de panne, car il est admis que le re-calcul d'un partage est plus rapide que le calcul d'une route alternative et devient un moyen d'implémenter la protection par commutation. Dans l'intégralité de cette thèse, nous faisons l'hypothèse qu'il est toujours possible de calculer un nombre donné de chemins entre deux points du réseau. Cette hypothèse est essentielle pour permettre aux opérateurs d'optimiser leurs réseaux au moyen d'une répartition et d'une distribution du trafic et, en même temps, d'être capable d'assurer un certain dégré de fiabilité

et de tolérance aux pannes.

4.10.2. Partage de charge et Routage IP classique

Les extensions pour le routage et l'acheminement multichemin du protocole IP permettent aux routeurs chargés d'acheminer un paquet de gérer un nombre arbitraire de «routeurs suivants» (next hops ou NH) pour une même destination, parmi lesquels le mécanisme de partage de charge doit en choisir un. Nous allons voir que la mise en oeuvre de ces extensions (nous allons nous limiter au contexte unicast) pose un certain nombre de problèmes pouvant limiter leur applicabilité. Pour une analyse détaillée, voir [89].

Une implémentation possible pour l'acheminement multi-chemin unicast consiste à utiliser des mécanismes de type «Round Robin» pour choisir le NH parmi les candidats possibles. Cette approche présente certains d'inconvénients :

1. *Variable Path MTU.* Les mécanismes de découverte de MTU *(Maximum Transfer Unit)* ne fonctionnent plus si les différents chemins ont des MTU différents.
2. *Variable Latency.* L'emprunt de routes différentes peut entraîner la réception non-ordonnée de paquets, pouvant occasionner un forte dégradation des performances de TCP.
3. *Métrologie.* L'acheminement multi-chemin peut rendre des outils comme «ping», «traceroute», ou d'autres plus évolués, inefficaces.

La plupart des problèmes évoqués sont dûs au fait que des paquets appartenant au même flot sont susceptibles d'emprunter des chemins différents. Ainsi, une solution possible consiste à assurer que deux paquets appartenant au même flot seront acheminés sur la même route physique.

Solutions Possibles

- *Modulo-N chemin* Le choix du NH parmi une liste de N possibilités est réalisé en évaluant une fonction de hachage modulo N sur les champs de l'en-tête permettant d'identifier un flot. Le principal inconvénient est que l'ajout ou la suppression de nouveaux candidats dans la liste des NH peut affecter l'acheminement des flots existants.
- *Hash-Threshold* L'espace d'arrivée de la fonction de hachage est divisé en intervalles, et chaque intervalle est attribué à un candidat. Avec cette approche il est possible de mieux contrôler l'effet sur les flots existants de l'ajout ou de la suppression de candidats.
- *Highest Random Weight (HRW)* La fonction de hachage prend en compte non seulement les champs de l'en-tête permettant d'identifier un flot, mais aussi l'adresse des NH candidats. Ici aussi, l'avantage de cette approche est une minimisation du nombre de flots affectés par une modification de la liste des candidats, mais est plus coûteuse en puissance de calcul.

Ainsi, la plupart des mécanismes proposés s'appuient sur des fonctions de hachage, et peuvent obtenir une distribution équitable si le nombre de flots est élevé. De ce fait, leur application est limitée au contexte des réseaux d'opérateur.

4.10.3. Taxonomie du partage de charge

Les différentes heuristiques, algorithmes, etc. *(ang. Schemes)* permettant de mettre en place le partage de charge peuvent être classés de plusieurs manières : on peut parler du partage de charge statique, dynamique, adaptatif, etc. En ce qui nous concerne, nous utiliserons la classification suivante : partage de charge orienté connexion et partage de charge orienté paquet.

Partage de charge orienté connexion

Dans les méthodes orientées connexion, les flots de données sont caractérisés par un nombre (idéalement réduit) de paramètres, et les décisions de routage et (ou) d'acheminement affectent le flot dans son ensemble. A ce sujet, on remarquera l'analogie avec le routage sensible à la Qualité de Service classique.

Partage de charge orienté paquet

Dans les méthodes orientées paquet, les décisions d'acheminement sont prises concernant chaque paquet individuellement. Les méthodes orientées paquet s'avèrent bien adaptées dans les réseaux à commutation de paquets non orientés connexion tel que les réseaux IP.

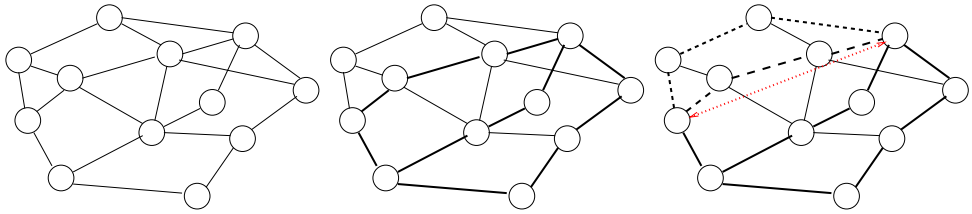
Méthodes hybrides

Il est généralement admis que le déséquencement des paquets n'est pas souhaitable (elle affecte en particulier la performance de protocole tels que TCP). Ceci justifie la nécessité de mécanismes orientés paquet dits hybrides capables de garantir que des paquets appartenant à une même connexion ou «micro-flot» sont acheminés sur un même LSP. Comme évoqué précédemment, une approche classique consiste à utiliser des tables de hachage. Rappelons que la quantité de connexions constituant le trafic agrégé doit être suffisamment élevée pour assurer la granularité des mécanismes de hachage. Nous considérons que cette hypothèse est respectée dans le contexte d'un réseau d'opérateur.

Dans un réseau sans classes de trafic, le problème du partage de charge peut se réduire, de façon simpliste, au calcul d'un vecteur de coefficients, un *partage,* donnant pour chaque LSP la proportion de trafic à acheminer. Si le modèle de service prend en compte plusieurs classes de service (par exemple une architecture à Différentiation de Services), on peut parler du calcul d'une matrice de proportions (par LSP et par classe). Si l'on établit un LSP par classe on peut réduire le prbolème au cas précédent.

4.10.4. Partage de charge MPLS et Formulation du problème

Considérons un ensemble de LSPs, issus d'une *route explicite* ayant les mêmes points terminaux. Nous appelons cet ensemble un *groupe* (de LSPs). Le groupe peut être établi manuellement ou dynamiquement.



(a) (b) (c)

Fig. 4.3: Groupe de LSPs. Le groupe est considéré comme une donnée du problème. Le partage de charge entre des chemins non disjoints peut être pris en compte par le partage de charge hiérarchique : un noeud du réseau est responsable d'agréger les informations de routage et d'annoncer une seule adjacence au routeur d'entrée (ingress).

4.10.4.1. Le calcul initial du groupe des LSPs

Chemins Disjoints

Si la topologie du réseau et d'autres contraintes de caractère général le permettent, le groupe de LSPs devrait être à liens disjoints, afin d'assurer un meilleur degré de tolérance aux pannes. Si nécessaire, le groupe peut être à noeuds disjoints (un groupe à noeuds disjoints implique un groupe à liens disjoints). Le calcul du groupe peut être réalisée «on line» ou «off line», utilisant des versions modifiées des algorithmes de «max-flow» ou «des K plus courts chemins» [26]. Ces algorithmes peuvent prendre en charge le marquage des éléments du réseau (nœuds ou liens) en utilisant un «degré» qui limite le nombre de LSPs différents les traversant.

Hypothèse

Dans la suite, nous faisons l'hypothèse que le groupe de LSPs est fixe, et donc que son optimisation est faite sur une échelle de temps plus grande que les mécanismes de partage de charge que nous proposons. Le groupe des LSPs est supposé stable, sauf dans le cas où des événements tels que des liens tombant en panne ou des mises à jour du routage imposent un re-calcul du groupe. Finalement, notons que le groupe peut être également le résultat d'une optimisation prenant en compte des contraintes administratives, telles que l'interdiction de traverser un certain Système Autonome. Le groupe peut varier (lentement) dans le temps et être communiqué aux routeurs concernés au moyen d'un protocole tel que COPS ou SNMP.

Notre problème du partage de charge peut être énoncé de la façon suivante : étant donné un groupe de LSPs [4] entre un nœud A et un nœud B, et étant donnée une caractérisation du trafic (la caractérisation utilisant des processus stochastiques étant la plus utilisée), déterminer la proportion relative du trafic qui doit être acheminée sur chaque LSP du groupe. L'objectif d'un tel partage est d'optimiser le réseau par rapport à un certain critère.

[4] Rappelons que les LSPs sont unidirectionnels.

4.10.5. Travaux existants sur le partage de charge

Le partage de charge dans les systèmes distribués a été étudié par différents auteurs dans différents contextes. Kleinrock [55] et Bertsekas et Gallager [10] ont formalisé le problème du routage optimal (nous préférons parler de répartition optimale de trafic sur un nombre pré-établi de chemins plutôt que de routage), en définissant un problème d'optimisation globale et en utilisant des fonctions de coût additives et convexes, choisies génériquement et approximant le délai moyen d'une file M/M/1. Dans ces travaux, le problème du partage de charge est formulé et résolu à l'aide d'une approche déterministe : à partir de la donnée de la matrice de trafic, (définissant des débits constants ou trafic Poissonien) et des fonctions de coût, les auteurs proposent un problème d'optimisation Lagrangienne sous contraintes. Ils obtiennent finalement des conditions d'optimalité. Ces conditions d'optimalité font intervenir les fonctions de coût et leurs dérivées. D'autres auteurs [102] ont évalué la complexité de cette approche. [46] propose un algorithme de partage de charge dans les réseaux IP en utilisant des technologies basées sur des agents.

Dans le contexte particulier des réseaux MPLS, dans [24] les auteurs proposent la répartition optimale du trafic sous l'hypothèse de trafic Poissonien en utilisant des résultats classiques de la théorie des files d'attente comme les réseaux de Jackson. [65] définit un système dynamique de partage de charge avec établissement et relâchement dynamique de LSPs, en tirant partie des informations fournies par un protocole de routage sensible à la qualité de service et en évitant les routes fortement chargées. Apparemment, cette approche a été refusée par l'IETF [5] du fait de sa complexité. Dans [62] les auteurs proposent une approche adaptative des résultats théoriques de Bertsekas et Gallager en estimant les dérivées des fonctions de coût (en l'occurence, des délais marginaux) à l'aide de paquets de test *(ang. packet probes)*.

4.10.6. Aspects architecturaux du partage de charge

Comme évoqué précédemment, dans [87] les auteurs proposent des architectures centralisées pour le calcul des routes et l'implémentation de l'ingénierie de trafic. Dans ces architectures, la notion de *Réseaux gérés par politiques* [6] joue un rôle fondamental, et des protocoles comme COPS ou SNMP sont souvent proposés comme moyens d'échange des requêtes et des décisions entre les différentes entités (LSRs). L'approche intégrant COPS dans un domaine MPLS prévoit des «Points d'application des politiques» *(PEP ou Policy Enforcement Point)* au niveau de chaque LSR (ou, au moins, de chaque E-LSR) et un ou plusieurs «Points de prise de décision de politiques» *(PDP ou Policy Decision Point)*. Cette architecture est illustrée sur la figure 4.4 : le routeur d'entré (son PEP) applique le partage de charge calculé par le PDP en utilisant les deux LSPs disponibles.

4.10.7. Le partage de charge MPLS et sa normalisation

Le groupe de travail MPLS-WG de l'IETF a rapidement identifié le partage de charge comme étant une application intéressante de l'architecture MPLS et de l'ingénierie de trafic. Un certain nombre de documents préparatoires *(drafts)* ont été publiés recemment [7]. L'architecture MPLS [91] et les extensions pour l'architecture à différentiation de services [92] laissent ouverte la possibilité que des

[5] Conversation privée avec M. Marco Carugi, membre du groupe de travail.
[6] Libre traduction du terme *Policy Managed Networks*
[7] http ://www.ietf.org/internet-drafts/draft-allan-mpls-loadbal-00.txt

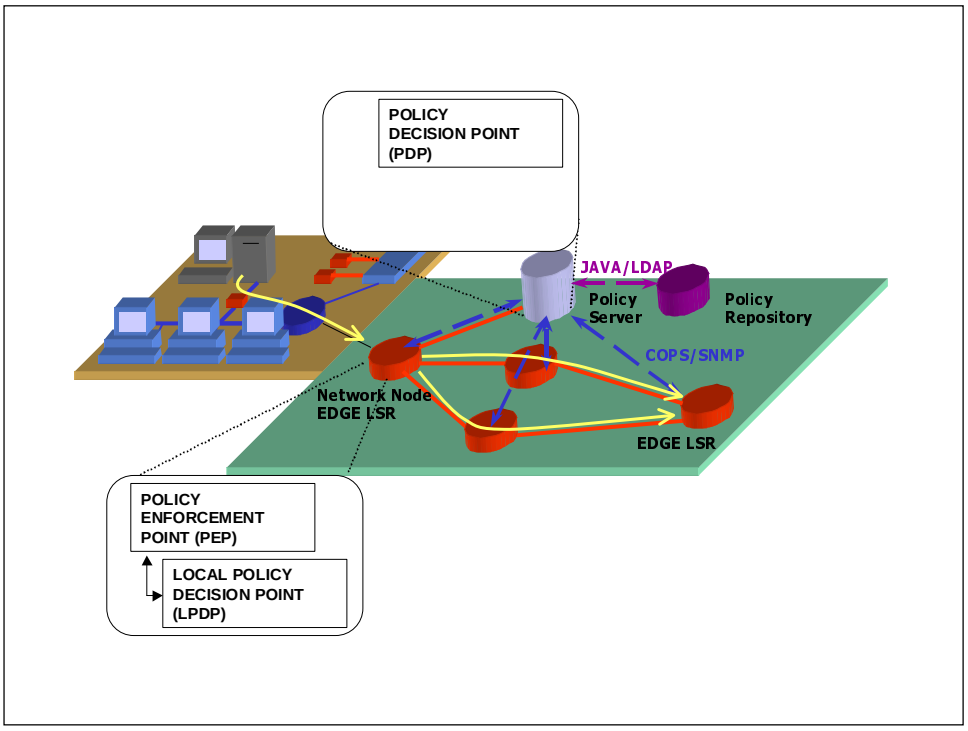


Fig. 4.4: Intégration du Partage de charge dans une solution COPS

instances individuelles des tables FTN (*FEC TO NHLFE*) et ILM (*Incoming Label Map*) pointent vers plusieurs entrées dans la table NHLFE *(Nex Hop Label Forwarding Entry),* en garantissant l'unicité de l'élément choisi. La procédure de sélection de l'entrée dans la table NHLFE indiquant le routeur suivant et l'étiquette de sortie n'est pas spécifiée. Cette ouverture permet d'envisager un partage de charge entre plusieurs LSPs différents entre deux points du réseau. Néanmoins, les mécanismes de sélection d'entrée doivent respecter certaines contraintes :

Unicité Qu'un élément et un seul ne soit choisi.

Ré-ordonnancement minimal de paquets d'un même flot Cette contrainte est respectée en s'assurant que des paquets appartenant à un même flot sélectionnent la même NHLFE.

Equité («Fairness») et famine Concernant la distribution du trafic entre l'ensemble de NHLFEs candidats, les normes ne spécifient rien quant au choix des NHLFE (par round robin, mécanismes de hachage etc.), mais insistent sur la nécessité d'une équité.

Cohérence Le trafic de «test», associé à un flot (agrégat ou non) doit utiliser la même entrée NHLFE que le trafic usager, car dans le cas contraire, les mécanismes mis en place pour détecter ou diagnostiquer de façon pro-active (préventive) des pannes éventuelles ou la dégradation de la

qualité de service, peuvent donner des résultats incohérents. Autrement dit, les procédures destinées à surveiller le trafic à l'aide de trafic de test doivent prendre en compte le partage de charge.

Préservation des caractéristiques DiffServ Le fait d'implémenter le partage de charge ne doit pas modifier les caractéristiques concernant la différentiation du trafic.

4.11. Conclusions

Dans ce chapitre, nous avons présenté les composantes relatives à l'ingénierie de trafic du plan de contrôle MPLS, et leurs relations. Nous avons insisté sur la notion de partage de charge, en présentant ses avantages notables et comment celui-ci peut être implementé de façon relativement simple dans les réseaux MPLS.

Deuxième partie .

Optimisation et Dimensionnement

Avant Propos

Les quelques citations suivantes illustrent notre approche.

L'ingénieur réseau est toujours confronté à la complexité. Les systèmes sont complexes, les interactions entre éléments sont complexes, les effets de bord sont complexes... Si les outils qu'il utilise pour mener à bon terme ses objectifs vis-à-vis des réseaux sont également complexes, ces outils font plus partie du problème que de la solution[8].

Even if you could afford it, over provisioning doesn't satisfy all application service-level requirements [...]. Each application requires different service from the network and behaves in different ways. Even generous over provisioning, in fact, may not provide the assurance that the network will handle the specific demands of certain applications adequately. [9]

The best way to get answers is to just keep working [at] the problem, recognizing when you are stalled, and directing the search pattern.[...] Don't just wait for The Right Thing to strike you, try everything you think might even be in the right direction, so you can collect clues about the nature of the problem. [10]

Il n'existe pas une approche unique pour la résolution d'un problème : les approches mathématiques fournissant des résultats analytiques sont basées sur la conception d'un modèle qui impose parfois des hypothèses non réalistes ou qui ne prend pas en compte toutes les interactions ayant lieu dans le système réel que l'on essaie de modéliser, d'évaluer ou tout simplement de comprendre. Les approches basées sur des simulations peuvent nous aider à mieux comprendre ces interactions, et elles constituent normalement l'étape préalable à la proposition d'un modèle mathématique. Pourtant, les simulations présentent le risque de ne pas prendre en compte tous les cas possibles, de négliger certains cas pathologiques importants et d'être limitées par les outils sous-jacents. Les conclusions tirées de certains scénarii (utilisant, par exemple, des topologies réseau simplistes) peuvent ne pas être généralisables à tous les cas envisagés. Même si un modèle analytique générique est conçu ou si l'on envisage des simulations exhaustives, on rencontre toujours les problèmes liés à la complexité, le temps ou la puissance de calcul nécessaire. La solution fait souvent appel à l'expérience, à une connaissance approfondie du système dans sa globalité et à une maîtrise des outils disponibles.

Dans cette deuxième partie, nous présentons le noyau théorique de cette thèse. Le lecteur est supposé avoir une connaissance de base de la théorie des probabilités, notamment des notions de variables aléatoires et de processus stochastiques, ainsi que les éléments essentiels de la théorie des grandes déviations. Pour le lecteur non familier avec cette théorie, nous conseillons la lecture de l'annexe

[8] Libre adaptation de Hoare - 1980 ACM Turing Award Lecture.
[9] Cisco Packet Magazine, http ://www.cisco.com/warp/public/784/packet/oct00/p62-cover.html, Oct. 2000
[10] John Carmack, id Software, Août 2002.

A, *Techniques des Grandes Déviations* dans laquelle nous présentons les éléments essentiels de cette théorie dans le contexte de l'évaluation de performances des réseaux de files d'attente.

Les travaux présentés dans cette partie ont été développés dans le contexte technologique présenté dans la première partie. Néanmoins, certains résultats sont suffisamment génériques pour être applicables non seulement à des réseaux à commutation de paquets autres que MPLS, mais également à d'autres domaines dans lesquels les systèmes physiques peuvent être modélisés de façon similaire aux modèles que nous développons dans la suite, comme le contrôle automatique.

Quidquid latine dictum sit, altum sonatur.

5. Outils Mathématiques

5.1. Motivation

Comme évoqué dans l'introduction, un des objectifs de cette thèse est d'optimiser et évaluer des mécanismes de partage de charge et d'ingénierie de trafic en proposant des modèles analytiques génériques. La caractérisation utilisé des processus stochastiques modélisant les processus des arrivées de paquets dans ces modèles est un critère important pour déterminer l'applicabilité d'un modèle et à cet égard, nous avons considéré la théorie des bandes passantes effectives qui nous permet de prendre en charge une famille importante de processus.

5.2. Introduction

La caractérisation du trafic a toujours été un sujet d'importance capitale dans le contexte de la modélisation et de l'évaluation de performances des réseaux. Dans un premier temps, des modèles utilisant des processus de Poisson ont été utilisés, notamment dans le contexte de l'optimisation des réseaux téléphoniques. Ce type de processus modélise bien les arrivées des appels téléphoniques classiques, et sont relativement simples : ce sont des processus à accroissements indépendants, et ils présentent certaines propriétés intéressantes comme PASTA (Poisson Arrivals See Time Averages). La forte expertise acquise justifiait son application aux réseaux de paquets.

Néanmoins, on sait que l'utilisation de processus de Poisson pour la modélisation des arrivées de paquets dans un réseau à commutation de paquets n'est pas réaliste. Un certain nombre d'études ont mis en évidence les limites des modèles théoriques utilisés historiquement et montré que le trafic présente des propriétés de mémoire longue et d'auto-similarité (notions définies dans la suite). La communauté scientifique n'a cessé de développer de nouveaux modèles de trafic, utilisant des processus stochastiques plus complexes, ou développant et appliquant de nouvelles théories mathématiques. Dans ce cadre, la théorie des bandes passantes effectives (b.p.e.) fournit une caractérisation élégante, utile et relativement complète des processus modélisant le trafic.

Les travaux de cette thèse sont développés autour de cette théorie. L'objectif de ce chapitre est d'en présenter les éléments les plus importants. Néanmoins, on ne peut pas découpler la théorie des bandes passantes effectives de celle des grandes déviations mais, pour alléger l'exposé, nous donnons ici la définition de bande passante effective, ainsi que quelques propriétés des bandes passantes effectives et ses applications directes aux réseaux de files d'attente. Les résultats des grandes déviations sous-jacents sont détaillés en annexe (cf. annexe A, *Techniques des Grandes Déviations*). Dans la section 5.3 nous étudions le problème du contrôle d'admission en utilisant une approche classique ce qui nous permet de mettre en évidence certaines limitations d'une telle approche tout en présentant quelques définitions qui seront reprises après avoir défini la notion de bande passante effective. La section 5.4 définit quelques notions de la théorie du télé-trafic qui nous seront utiles dans la suite. Ensuite, nous définissons la notion de bande passante effective dans la section 5.5. Nous donnons l'expression de la

bande passante effective pour quelques modèles de trafic simples dans la section 5.6, nous parlons de façon succincte de l'estimation de bandes passantes effectives dans la section 5.7 et dans la section 5.8 nous formalisons le lien avec la théorie des grandes déviations et les applications connues sous le nom de *Asymptotique du Grand Buffer* et *Asymptotique du Grand Nombre d'Usagers.*

5.3. Le problème du contrôle d'admission

Le problème du contrôle d'admission, et de manière générale, le problème du contrôle et évaluation de performances du partage d'une ressource entre plusieurs sources concurrentes a motivé le développement de la théorie des bandes passantes effectives.

A ce propos, avant de présenter la notion de bande passante effective, nous étudions brièvement un problème classique adapté à notre contexte, le problème du contrôle d'admission, qui consiste à déterminer si une nouvelle connexion peut être acceptée en respectant le niveau de qualité de service promis à toutes les sources acceptées. Considérons un ensemble de classes d'usagers $\{1..I\}$. A classe d'usagers appartiennent un nombre fini de sources, noté m_i pour la classe i. Toutes les sources d'une même classe ont un même débit de transfert, que nous supposons constant et que nous notons $r_i$pour la classe i.

L'administrateur réseau dispose d'un lien de transmission de capacité C. Nous faisons également l'hypothèse que le fait d'accepter une source de classe i rapporte à l'administrateur un bénéfice b_i. L'objectif de l'administrateur est de maximiser son revenu en choisissant un vecteur $\mathbf{n} \triangleq (n_1, \ldots, n_I), n_i \in \mathbb{N}$ où n_i est le nombre de sources de classe i admises, tout en garantissant un système sans dépassement de la capacité. Ce problème peut être formulé de la façon suivante :

$$\max_{\mathbf{n}} \sum_{i=1}^{I} n_i b_i \tag{5.1}$$

avec les contraintes :

$$\begin{aligned} \sum_{i=1}^{I} n_i r_i &\leq C \\ 0 \leq n_i &\leq m_i \quad \forall i \end{aligned} \tag{5.2}$$

Ce problème correspond à celui du remplissage du «sac à dos» *(ang. bounded knapsack problem)* car on a fait l'hypothèse que le nombre de sources de chaque classe est connu et fini. Ce problème peut être résolu à l'aide de la programmation linéaire ou des méthodes «branch and bound». Nous ne discutons pas ici de la complexité de ce type de problèmes, mais présentons l'analogie avec la notion de bande passante effective, qui sera formalisée dans la Section 5.5. Notons les points suivants :

- *Objectif et critère de performance :* il s'agit de maximiser le revenu total connaissant la capacité disponible.
- *Contraintes et contrôle d'admission :* ne pas dépasser la capacité du lien de transmission, contrôlant le partage de la ressource entre les sources concurrentes.
- *Région Admissible (ang. acceptance region) :* l'ensemble des combinaisons de sources respectant les contraintes.

– *Sous - optimalité :* l'ensemble des combinaisons de sources qui apportent un revenu $\geq K$ définit une région, sous-ensemble de la région admissible, qui dépend de K.
– *Combinaison linéaire et caractère additif :* le débit total offert au lien de transmission est obtenu comme combinaison linéaire des débits de chaque source : les coefficients de la combinaison linéaire correspondent aux composantes du vecteur **n**. Ceci est dû à la *propriété additive* de la bande passante.

Un tel dimensionnement s'avère relativement simple, principalement du fait que les valeurs r_i et b_i sont constantes. Hélas, une telle formalisation ne prend pas en compte les propriétés statistiques des sources. Nous allons voir que la théorie des bandes effectives permet un dimensionnement plus efficace par rapport à des approches classiques : d'une part, les approches basées sur le «pire cas» (ne considérant que les débits crêtes, par exemple) mènent à une mauvaise utilisation des ressources et un tel dimensionnement néglige l'effet du multiplexage statistique. D'autre part, les approches «simplistes», considérant uniquement les débits moyens peuvent ne pas être adaptées : à débit moyen égal, la sporadicité des sources peut avoir un impact très différent sur les critères de performance envisagés, comme par exemple le taux de pertes ou la probabilité de débordement d'une file d'attente.

5.4. Quelques définitions nécessaires

Dans cette section, nous donnons les définitions des principaux concepts dont nous aurons besoin par la suite.

Définition 5.4.1 (Processus stationnaire). Un processus stochastique $x(t)$ est dit **stationnaire** si $\forall n \in \mathbb{N}, \forall \tau, \forall t_0 < t_1 < t_2 < \ldots < t_n$ on a :

$$(x(t_0), \ldots, x(t_n)) \quad =_{\mathbb{L}} \quad (x(t_0+\tau), \ldots, x(t_n+\tau))$$

ou $=_{\mathbb{L}}$ indique égalité en loi.

On définit la fonction d'autocorrélation comme

$$\rho(\tau) = \frac{\mathrm{cov}[x(t), x(t+\tau)]}{\sqrt{\mathrm{var}[x(t+\tau)]\mathrm{var}[x(t)]}}$$

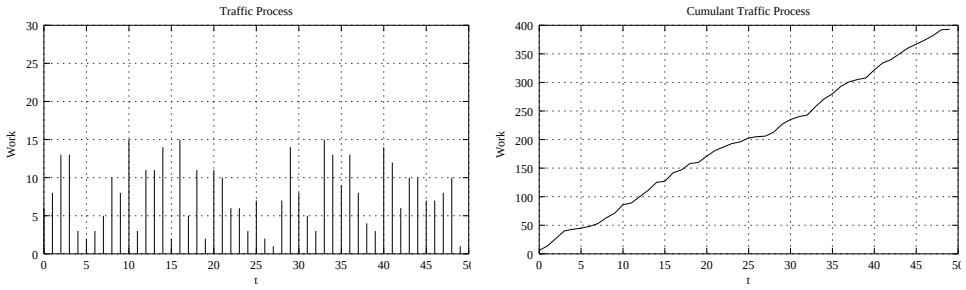
Le processus est dit **à covariance stationnaire** si

– $\mathbb{E}[x(t)] = \lambda < \infty$
– $\mathbb{E}[(x(t)-\lambda)^2] = \sigma^2 < \infty$
– $\mathbb{E}[(x(t)-\lambda)(x(t+\tau)-\lambda)] = \mathrm{cov}(\tau) < \infty$

Un processus stationnaire dont les deux premiers moments sont finis est à covariance stationnaire. Pour un processus à covariance stationnaire, la fonction d'autocorrélation est de la forme suivante :

$$\rho(\tau) = \frac{\mathrm{cov}(\tau)}{\sigma^2}$$

Définition 5.4.2 (Processus Cumulatif (ang. Cumulant Process)). Soit $x(t)$ un processus stochastique discret (resp. continu), et $t_0, t_1 \in \mathbb{N}$ (resp. $t_0, t_1 \in \mathbb{R}$). Le processus $X[t_0, t_1) \triangleq \sum_{t_0}^{t_1} x(t)$ (resp. $X[t_0, t_1) \triangleq \int_{t_0}^{t_1} x(t)dt$) est dit processus cumulatif (ou **processus d'accroissements**) de $x(t)$. (cf. figure 5.1).



(a) Processus instantané (b) Processus d'accroissements

Fig. 5.1: (a) Processus Stochastique à temps discret modélisant une source. Le travail produit par la source à chaque intervalle de temps est une v.a. de loi uniforme [1,15]. (b) Processus d'accroissements associé.

Définition 5.4.3 (Processus à accroissements indépendants). Un processus $x(t)$ est dit **à accroissements indépendants** si pour n'importe quelle suite d'instants de temps $0 = t_0 < t_1 < t_2 < \ldots < t_n$, les accroissements du processus $x(t_n) - x(t_{n-1}), x(t_{n-1}), \ldots, x(t_1) - x(t_0)$ sont indépendants.

Définition 5.4.4 (Processus à borne stationnaire). Un processus d'accroissements $x(t)$ est borné stationnairement si $\forall h$

$$\lim_{a\to\infty} \sup_t \mathbb{P}\left\{x(t+h) - x(t) \geq a\right\} = 0 \tag{5.3}$$

Définition 5.4.5 (Processus à mémoire longue (ang. Long Range Dependent)). Un processus $x(t)$ stationnaire est dit «à mémoire longue» *(ang. Long Range Dependent)* si

$$\sum_{k=-\infty}^{\infty} |\rho_x(k)| = \infty \tag{5.4}$$

Définition 5.4.6 (Modèles de Trafic à Queue Lourde). Une variable aléatoire X est dit «à queue lourde» si $\exists \alpha, 0 < \alpha < 2$ et $\exists C$ tel que $x^\alpha \mathbb{P}(|X| > x) \to C$, quand $x \to \infty$, où C est une constante et α est l'index de la distribution. Un processus avec des distributions marginales à queue lourde est dit un processus à queue lourde.

Définition 5.4.7 (Auto-similarité). Un processus $x(t)$ est dit «auto similaire» *(self-similar)* de paramètre H, si le processus $c^{-H}x(ct)$ et le processus $x(t)$ sont équivalents en distribution. L'exemple classique de processus auto similaire est le processus mouvement fractionnaire Brownien (fBm). Voir par exemple [96] pp. 34 ou [73].

5.5. La Bande Passante Effective

D'une façon générale, la notion de bande passante effective décrit la quantité de ressources qu'il est nécessaire d'allouer à un ensemble de sources afin de garantir un certain critère de performance.

Le terme «ressource» est souvent assimilé à l'idée de capacité d'un lien de transmission et celui de «critère de performance» à la probabilité de certains événements non souhaitables, ou à des métriques comme les délais moyens. Le caractère additif est souvent associé à l'idée de bande passante effective, par analogie au cas à débit constant, de sorte que la bande passante effective d'un agrégat de sources indépendantes est la somme de leurs bandes passantes effectives.

La notion de bande passante effective (certains auteurs francophones utilisent le terme «bande passante équivalente») est assez récente. Les premiers travaux de Hui [47] définissaient la notion de «envelope rate», similaire mais moins complète que la notion de bande passante effective que nous verrons par la suite. La théorie des bandes passantes effectives a été développée en s'appuyant sur celle des grandes déviations, théorie qui concerne l'étude et l'estimation de la probabilité de certains événements considérés comme «rares» ou dans certains régimes asymptotiques.

Définition de Bande Passante Effective

La *Bande Passante Effective (b.p.e.)* d'un processus d'accroissements ergodique à borne stationnaire est définie par [53] :

$$\alpha(s,t) = \sup_{t_0} \left\{ \frac{1}{st} \log \mathbb{E} \left\{ e^{s(X(t_0+t)-X(t_0))} \right\} \right\} \tag{5.5}$$

Si la source est stationnaire, le travail produit par la source pendant l'intervalle de temps $(t_0, t_0 + t)$,$X(t_0 + t) - X(t_0)$ est une variable aléatoire, notée $X(0, t]$ dépendant uniquement de la durée de l'intervalle. La *Bande Passante Effective* [1] devient :

$$\alpha(s,t) = \frac{1}{st} \log \mathbb{E} \left\{ e^{sX(0,t)} \right\} \tag{5.6}$$

Remarques

Pour des raisons historiques, la notion de bande passante effective est souvent mal interprétée et utilisé par abus de langage :

1. La notion de b.p.e. d'une source correspond à la définition (eq. 5.6) qui fait intervenir la transformée de Log-Laplace d'une source. A cet égard, la bande passante effective dépend uniquement des propriétés stochastiques des processus d'accroissements modélisant les arrivés au système, et est une caractérisation formelle intrinsèque à la source. La b.p.e. est définie comme une fonction de deux paramètres, un paramètre spatial noté s et un paramètre temporel noté t. C'est la connaissance de ces deux paramètres, déterminant un «point de travail» noté (s^*, t^*), qui détermine **quantitativement** la valeur de la bande passante effective de la source. La façon exacte de déterminer le point de travail dépendra du critère choisi et sera définie ultérieurement.
2. La formalisation que nous utiliserons du problème d'allocation de ressources entre différentes sources appartenant à un certain nombre de *classes distinctes* afin de garantir un certain critère de performance fait intervenir les b.p.e. des différentes sources. Sous certaines hypothèses (par exemple, si les sources son indépendantes) cette formalisation s'exprime en termes de

[1] Certains auteurs parlent aussi de *Scaled Logarithmic Moment Generating Function* comme transformation de la transformée de Laplace de la variable aléatoire.

combinaisons linéaires de fonctions caractérisant chaque classe (typiquement, leurs fonctions génératrices c-à-d leurs *bandes passantes effectives*) dont les coefficients représentent le nombre ou la proportion de sources de chaque classe. On retrouve à nouveau ici la notion d'additivité, qui explique le choix de cette nomenclature par analogie au problème de contrôle d'admission classique.

3. Si l'on définit la notion de «contexte de multiplexage» comme l'ensemble des ressources concernées (par exemple, la capacité du lien de transmission et la taille du buffer) et le nombre et type de sources concurrentes, c'est la connaissance du contexte de multiplexage qui permettra déterminer le point de travail du système et de calculer la quantité de ressource qui est allouée à chaque source afin de respecter le critère choisi. Autrement dit, la quantité de ressource attribuée à une source (l'évaluation de sa b.p.e. au point de travail) dépend de son contexte de multiplexage.
4. *En conséquence,* par abus du langage, la notion de bande passante effective présente une dualité : d'une part est une caractérisation formelle des propriétés stochastiques des sources (et donc intrinsèques à celles-ci) et d'autre part définit la quantité de ressources consommées / attribuées (et dépendantes donc de son contexte du multiplexage).

Remarquons également que la théorie des grandes déviations utilise typiquement la notation θ pour désigner le paramètre spatial. Nous utiliserons les deux notations indistinctement. L'intérêt de l'utilisation des bandes passantes effectives est le suivant :

1. Rappelons que la caractérisation d'une source par sa bande passante effective fait intervenir la fonction génératrice du processus de trafic associé, donc permettant une caractérisation complète, à quelques détails techniques près.
2. Les bandes passantes effectives interviennent directement dans les «principes de grandes déviations» appliqués à la théorie des files d'attente (ce terme sera défini ultérieurement). Ces principes permettent d'obtenir des approximations probabilistes de certains événements intéressants, et seront à la racine des différents critères de performance utilisés dans nos travaux.

Lors des premiers travaux sur les bandes passantes effectives, cette notion avait une application «locale» limitée à une seule file d'attente. Actuellement, certains travaux essaient de caractériser la variation de la b.p.e. par son passage par une file d'attente, afin de pouvoir étendre son application à des réseaux de files d'attente et d'obtenir des règles de dimensionnement dans un contexte globale de bout en bout.

Même si elle est générique, la théorie des bandes passantes effectives présente plusieurs inconvénients, dont un certain nombre sont détaillés dans cette thèse. Il est parfois difficile de donner une formule analytique simple de la bande passante effective d'une source ; certains processus présentent des bandes passantes effectives faisant intervenir des exponentielles de matrices et généralement, les estimateurs et les calculs faisant intervenir les bandes passantes effectives nécessitent des traitements numériques.

5.5.1. Propriétés des Bandes Passantes Effectives

Nous présentons ici quelques propriétés importantes des b.p.e. Pour une liste exhaustive, le lecteur est invité à consulter Kelly [53] ou Hui [47].

– La transformée de Log-Laplace d'une v.a. est convexe (cf. annexe). Cette propriété nous permettra d'assurer l'unicité du point de travail spatial (s^*).
– La bande passante effective associée à un agrégat de sources indépendantes est la somme de leurs bandes passantes effectives. Nous retrouvons ici la propriété d'additivité que nous avons évoquée précédemment. Cette additivité provient du fait que l'espérance du produit de deux v.a. indépendantes est le produit de leurs espérances, et du fait que le logarithme d'un produit est la somme des logarithmes.
– Si le processus de trafic est à accroissements indépendants, sa bande passante effective ne dépend pas de t.
– Pour un t fixé, la b.p.e. est croissante avec s, et on peut montrer que :

$$\frac{\mathbb{E}\left[X\left(0,t\right)\right]}{t} \leqslant \alpha\left(s,t\right) \leqslant \frac{\bar{X}\left(0,t\right)}{t} \tag{5.7}$$

La majoration correspond au *supremum essentiel,* défini par :

$$\bar{X}\left(0,t\right) \triangleq \operatorname{esssup}\left(X\left(0,t\right)\right) = \sup\left\{x : \mathbb{P}\left\{X\left(0,t\right) > x\right\} > 0\right\} \tag{5.8}$$

En pratique, cela veut dire que la valeur quantitative de la bande passante d'une source se trouve toujours entre sa le débit moyen et le débit crête de la source.

Interprétation

Il est difficile de donner une interprétation physique des paramètres s et t, et cette interprétation dépend en effet du critère choisi qui permet de calculer le point de travail. Pour tout t, la bande passante effective varie entre le débit moyen du processus de trafic instantané et son débit crête. Le paramètre spatial s^* reflète cette idée : plus il est petit, plus la b.p.e. est proche du débit moyen, et plus il est grand, plus la b.p.e. est proche du débit crête. Formellement :

$$\begin{aligned} (s \to 0) \quad & \alpha\left(s,t\right) = \frac{1}{t}\mathbb{E}\left[X\left(0,t\right)\right] + \frac{s}{2t}\operatorname{Var}\left[X\left(0,t\right)\right] + o\left(s\right) \\ (s \to \infty) \quad & \alpha\left(s,t\right) = \frac{1}{t}\bar{X}\left(0,t\right) + \frac{1}{st}\log\mathbb{P}\left\{X\left(0,t\right) = \bar{X}\left(0,t\right)\right\} + o\left(\frac{1}{s}\right) \end{aligned} \tag{5.9}$$

De façon générale, le paramètre t détermine la fenêtre de temps à considérer, pendant la quelle la source émet du trafic.

5.6. Quelques modèles de trafic et leurs Bandes Passantes Effectives

5.6.1. Loi Discrète

Proposition 5.6.1 (Bande Passante Effective associée à une source discrète). *Soit X_i une suite de variables aléatoires discrètes i.i.d, prenant leurs valeurs dans l'espace $E = \{a_0, a_1, a_2, \ldots, a_M\}$, $a_i \in \mathbb{R}^+$. Soit $X_{Disc}(n)$ le processus défini par la suite X_i et $X[0,n)$ le processus d'accroissements correspondant. En utilisant la propriété d'indépendance, on voit que :*

$$\alpha(s,t) = \frac{1}{s}\log\left(\sum_{k=0}^{M} e^{sk}\mathbb{P}(X_i = k)\right)$$

5.6.2. Processus de Bernoulli

Un cas classique dans la théorie de b.p.e. est $M = 1, \mathbb{P}(X_i = 1) = p, \mathbb{P}(X_i = 0) = 1 - p$ (loi de Bernoulli). Soit $X_{\text{Bern}}(n)$ un processus à temps discret défini par une suite de variables i.i.d. $X_i \sim Bernoulli(p)$. La bande passante effective associée au processus est donnée par :

$$\alpha_{\text{Bern}}(s,t) = \frac{1}{s}\log\left(pe^s + 1 - p\right) \tag{5.10}$$

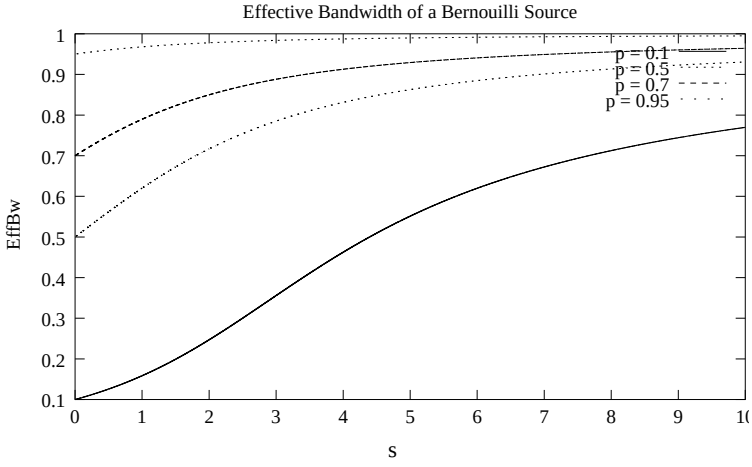


Fig. 5.2: Bande passante effective associée à $X_{\text{Bern}}(n)$, pour différents valeurs de p.

5.6.3. Processus de Poisson

Soit Y une v.a. de Poisson, c.-à-d. $Y \sim \text{Poisson}(\lambda)$. Alors :

$$\begin{aligned}\mathbb{E}\left[e^{(\theta Y)}\right] &= \sum_{n=0}^{\infty} e^{\theta n}\frac{\lambda^n}{n!}e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^\theta\lambda)^k}{k!} = e^{-\lambda}e^{e^\theta\lambda} \\ &= e^{\lambda(e^\theta - 1)}\end{aligned} \tag{5.11}$$

Pour un processus de Poisson, le nombre d'arrivées pendant l'intervalle de temps $(0,t]$ est une v.a. de Poisson de paramètre λt. Nous avons donc :

$$\begin{aligned}\alpha_{\text{Poisson}}(s,t) &= \frac{1}{st}\log\left(e^{\lambda t(e^s-1)}\right) \\ &= \frac{\lambda}{s}\left(e^s - 1\right)\end{aligned} \tag{5.12}$$

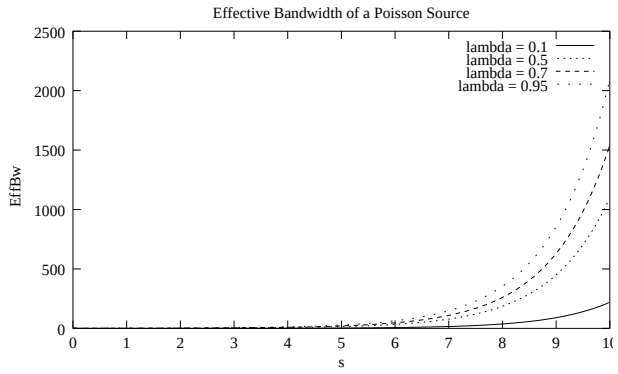


Fig. 5.3: Bande passante effective d'un processus de Poisson.

La dépendance du paramètre t n'apparaît pas dans les expressions 5.10 et 5.12, du fait de la propriété d'accroissements indépendants.

5.6.4. Mouvement Brownien Fractionnaire

Un autre exemple est le trafic modélisé par un mouvement Brownien fractionnaire [73], de moyenne λ de variance σ^2 et de paramètre de Hurst H. Sa bande passante effective est donnée par :

$$\alpha(s,t) = \lambda + \frac{1}{2}\sigma^2 s t^{2H-1}$$

Ce cas est analysé en détail dans la section 6.6.1.

5.7. Métrologie et Estimation des Bandes Passantes Effectives

Dans la section 7.6, nous proposons un mécanisme de partage de charge adaptatif basé sur la métrologie et l'estimation de bandes passantes effectives à partir de traces. Nous donnons ici quelques éléments importants. A ce sujet, à la base de cette estimation on trouve l'idée clé d'ergodicité.

Le nombre d'études portant sur l'estimation des bandes passantes effectives à partir de traces est relativement limité. Cette tâche s'avère encore plus difficile si le trafic est à mémoire longue, car l'analyse statistique nécessite alors pour une estimation correcte l'obtention d'un nombre important de traces indépendantes. Ceci pose un certain nombre de problèmes : la corrélation entre échantillons temporellement distants peut être importante dans le cas d'un trafic à mémoire longue. D'autre part, nous verrons que sous certains régimes asymptotiques, les événements intéressants tels que les débordements de capacité se succèdent dans des intervalles de temps pouvant beaucoup varier pour différents modèles de trafic, et ces intervalles de temps vont également définir la durée minimale des traces ; enfin, le calcul des intervalles de confiance dépend du modèle de trafic considéré. En résumé, sauf dans certains cas particuliers, l'estimation des bandes passantes effectives à partir de traces reste aujourd'hui un sujet complexe nécessitant des études approfondies.

Les travaux concernant l'estimation des bandes passantes effectives sont complémentaires à nos

études, et constituent un axe de recherche intéressant dans le contexte de projets de métrologie. Certaines méthodes ont été développées dans des régimes asymptotiques particuliers (l'asymptotique du grand buffer qui sera définie par la suite), et font intervenir des bandes passantes effectives dépendant du seul paramètre spatial. De plus, la connaissance «a priori» de la loi du processus d'arrivées peut amener à des meilleurs estimateurs, à l'aide de l'estimation paramétrique. Citons comme exemple l'estimation paramétrique de la b.p.e. d'un processus de Poisson, qui consiste à estimer sa moyenne λ et à en déduire $\hat{\alpha}_{\text{Poisson}}(s)$ grâce à l'équation 5.12.

L'estimateur de Dembo

Considérons une source de trafic générant $x[t]$ unités de travail à l'instant t. Amir Dembo ([96] pp.43) propose une méthode d'estimation de la transformée de Log-Laplace du processus d'accroissements de la source : on dispose de n traces d'une réalisation du processus. Considérons la taille de bloc b pour laquelle les sommes :

$$X_1 \triangleq \sum_{t=1}^{b} x[t], X_2 \triangleq \sum_{t=b+1}^{2b} x[t], \dots \tag{5.13}$$

sont i.i.d. L'estimateur de la transformée de Log Laplace du processus est alors donné par :

$$\frac{1}{b} \log \left(\frac{1}{\lfloor n/b \rfloor} \sum_{i=1}^{\lfloor n/b \rfloor} e^{\theta X_i} \right) \tag{5.14}$$

et, la b.p.e. peut être estimée par :

$$\hat{\alpha}(s,t) = \frac{1}{st} \log \left(\frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} e^{sX_i} \right) \tag{5.15}$$

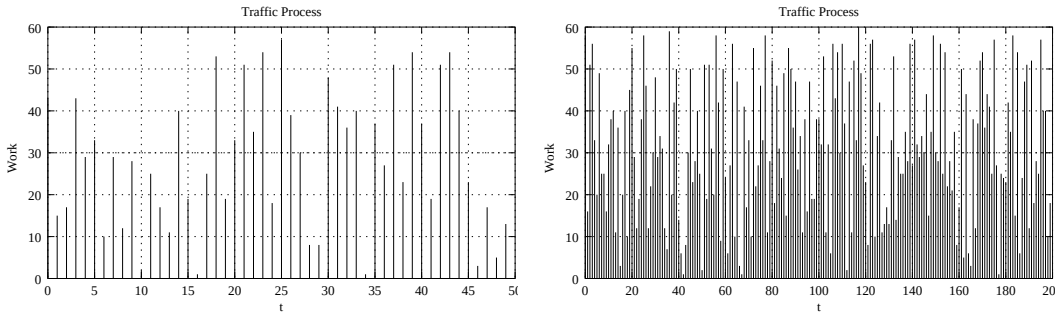
Cet estimateur est utilisé dans les mécanismes de partage de charge adaptatifs que nous proposons dans la section 7.6, où nous discutons de ses avantages et inconvénients.

Exemple

Afin d'illustrer l'estimation des b.p.e. à partir de traces *(ang. measurement-based effective bandwidth estimation),* les figures 5.4, 5.5 et 5.6 illustrent l'estimation de Dembo de la b.p.e. d'un processus à temps discret, comparé à la b.p.e. réelle. Ce procesus est généré par simulation, avec un point de travail $t^* = 1$: à chaque intervalle de temps, le travail généré par une source suit une distribution uniforme sur l'intervalle $[0, 60]$, et les intervalles sont indépendants. La figure 5.4 illustre une trajectoire pour $(a)N = 50, (b)N = 200$ échantillons. Nous verrons dans les chapitres suivants que cet intervalle de temps est important dans les systèmes sans buffer. De plus, cette estimation sera utilisée dans la deuxième partie du chapitre 7, *Partage de charge à Capacité variable*

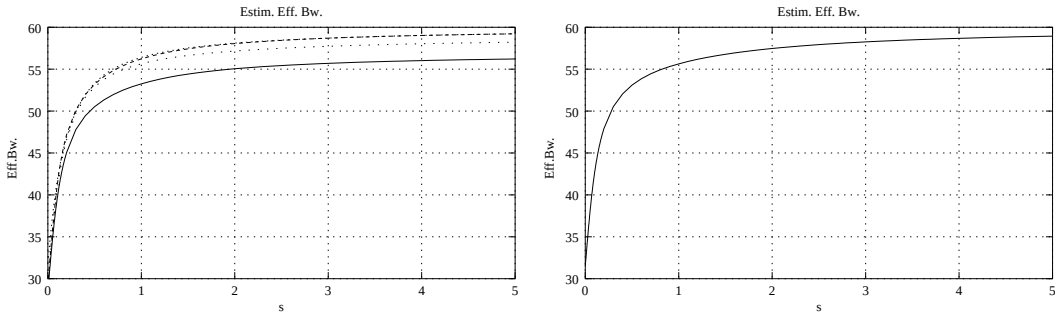
Autres méthodes : Rabinovitch [96] présente d'autres méthodes d'estimation de bandes passantes effectives, notamment des méthodes basées sur le ré-échantillonnage et des mécanismes de bootstrap et *surrogate data,* en utilisant des méthodes appelées de «blocs lissants». Sa principale contribution

est l'estimation des b.p.e. pour les flots de trafic présentant une dépendance longue, et pour lesquels l'application des mécanismes d'estimation de b.p.e. classiques est limitée. Finalement, il est intéressant de noter que pour l'estimation de certains événements, comme les probabilités de pertes ou des débordements, il n'est pas nécessaire d'estimer la bande passante effective de la source. Rabinovitch [96] propose pour cela une méthode graphique basée sur la simulation et les régressions linéaires, et utilisant certaines principes des grandes déviations.



(a) n=50 (b) n=200

Fig. 5.4: Trajectoires d'un processus de trafic discret, distribution uniforme

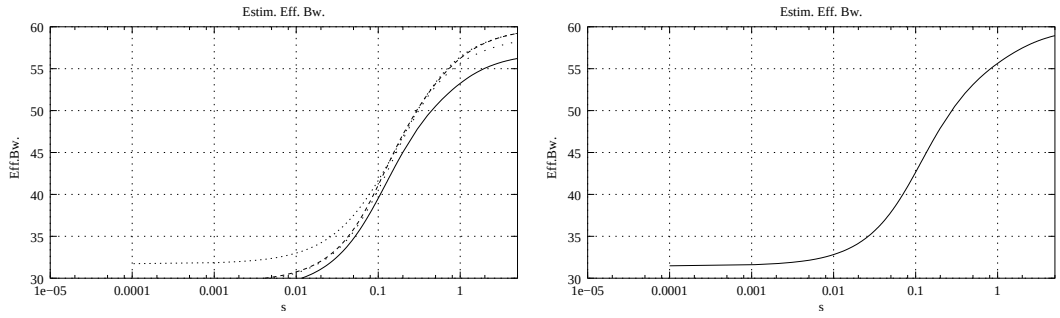


(a) n=50 (b) n=200

Fig. 5.5: Estimation Ergodique de la b.p.e.

Commentaires

D'une manière générale, l'estimation de bandes passantes effectives pour des valeurs arbitraires de s et t reste une tâche complexe. Comme les modèles de trafic différents présentent des échelles de temps caractéristiques différentes (nous verrons ceci dans la suite, lorsque nous parlerons de l'asymptotique de grand nombre d'usagers), on ne dispose pas d'éléments suffisants pour savoir a priori pendant quel intervalle de temps mesurer.



(a) n=50 (b) n=200

Fig. 5.6: Estimation Ergodique de la b.p.e.(échelle logarithmique)

5.8. Applications des b.p.e aux files d'attente

Dans cette section, nous présentons d'abord quelques résultats préliminaires : le premier concerne la transformation de mesures et le deuxième est connu comme le théorème de Bahadur-Rao [53]. Ces deux résultats sont nécessaires pour présenter les deux applications principales de la théorie des bandes passantes effectives aux files d'attente, connues comme *l'asymptotique de grand buffer* et *l'asymptotique de grand nombre d'usagers.*

5.8.1. Résultats Préliminaires

Transformations de Mesures (Cramér, cf. [21][23], pag.32)

Soit μ une distribution de probabilité sur $\mathbb{R}$, et soit $X \sim \mu(dx)$. Soit $L(\theta)$ (resp. $\Lambda(\theta)$), sa transformée de Laplace (resp. LogLaplace),

$$\begin{aligned} L(\theta) &= \mathbb{E}_\mu[e^{\theta X}] = \int_{\mathbb{R}} e^{\theta x}\mu(dx) \\ \Lambda(\theta) &= \log L(\theta) \end{aligned} \tag{5.16}$$

Soit $\Lambda^*(x)$, la transformée de Fenchel-Legendre (transformée convexe) de la transformée de LogLaplace de cette distribution de probabilité, définie par :

$$\Lambda^*(x) = \sup_\theta \{\theta x - \Lambda(\theta)\} \tag{5.17}$$

Notons $D = \{\theta : L(\theta) < \infty\}$, le domaine de la transformée de Laplace. On fait l'hypothèse que $\exists c \in \mathbb{R} t.q. \mu(\{c\}) \neq 1$ et que $\exists \theta_c^* \in D$, solution de l'équation $\frac{L^{'}(\theta)}{L(\theta)} = c$. Cette condition est équivalente à *trouver l'argsup de la transformée convexe (également appelée fonction de taux) évaluée en c.* Si

l'on définit la distribution de probabilité

$$\mu_c(dx) = \frac{e^{\theta^* x}}{L(\theta^*)}\mu(dx) \tag{5.18}$$

alors une v.a. X de distribution $\mu_c(dx)$ a tous ses moments finis et $\mathbb{E}_c[X] = c$ et $Var_c[X] = \frac{L^{''}(\theta^*)}{L(\theta^*)} - c^2 > 0$

Exemple 5.8.1 *(v.a. Gaussienne).* Afin d'illustrer cette notion, un cas simple particulièrement intéressant pour la suite est le suivant : Considérons une v.a $X \sim \mathcal{N}(p\lambda, p^2\sigma^2)$, avec $0 < p < 1$. Les transformées de (Log) Laplace de X sont données par :

$$\begin{aligned} L(\theta) &= \frac{1}{\sqrt{2\pi\sigma^2 p^2}} \int_{-\infty}^{\infty} e^{\theta x} e^{-\frac{1}{2}\frac{(x-\lambda p)}{\sigma^2 p^2}} dx \\ &= e^{\left(\theta\lambda p + \frac{1}{2}\theta^2\sigma^2 p^2\right)} \\ \Lambda(\theta) &= \theta\lambda p + \frac{1}{2}\theta^2\sigma^2 p^2 \end{aligned} \tag{5.19}$$

et la transformée de Fenchel-Legendre

$$\begin{aligned} \Lambda^*(c) &= \sup_{\theta}\left\{\theta c - \theta\lambda p - \frac{1}{2}\theta^2\sigma^2 p^2\right\} \\ \theta_c^* &= \frac{c-\lambda p}{p^2\sigma^2} \end{aligned} \tag{5.20}$$

et donc

$$\Lambda^*(c) = \frac{1}{2}\frac{(c-\lambda p)^2}{\sigma^2 p^2} \tag{5.21}$$

Finalement, notons que

$$\frac{e^{x\theta^*}}{L(\theta^*)} = e^{\frac{1}{2}\frac{(c-\lambda p)(2x-(c+\lambda p))}{p^2\sigma^2}} \tag{5.22}$$

et donc

$$\begin{aligned} \mathbb{E}_c[e^{\theta X}] &= \frac{1}{\sqrt{2\pi\sigma^2 p^2}} \int_{-\infty}^{\infty} e^{\theta x} \frac{e^{x\theta^*}}{L(\theta^*)} e^{-\frac{1}{2}\frac{(x-\lambda p)}{\sigma^2 p^2}} dx \\ &= e^{\left(c\theta + \frac{1}{2}\theta^2 p^2\sigma^2\right)} \\ \frac{\partial \mathbb{E}_c[e^{\theta X}]}{\partial\theta}\,|_{\theta=0} &= c \\ \frac{\partial^2 \mathbb{E}_c[e^{\theta X}]}{\partial\theta^2}\,|_{\theta=0} &= p^2\sigma^2 + c^2 \Rightarrow Var_c[X] = p^2\sigma^2 > 0 \end{aligned} \tag{5.23}$$

Théorème de Bahadur-Rao (cf. p.ex. [58])

Théorème 5.8.1 (Bahadur Rao). *Soit $L(\theta)$ (resp. $\Lambda(\theta)$) la transformée de Laplace (resp. LogLaplace) d'une v.a. X. Soit $(X_n|n \in \mathbb{N})$ une suite de v.a. i.i.d. de même loi que X. Supposons que $\Lambda(\theta)$ est finie sur $\mathbb{R}$, et que la loi de X est absolument continue. soit $S_n = \sum_{i=1}^{n} X_i$, alors :*

$$\mathbb{P}\left(S_n \geq nc\right) = \frac{1}{\theta_c\sqrt{2\pi n\sigma_c^2}} e^{-nI(c)}\big(1 + o(1)\big) \tag{5.24}$$

où

$$\begin{aligned} I(c) &= \sup_{\theta}\left\{\theta c - \Lambda(\theta)\right\} \\ \theta_c &= \text{argsup}\left\{\theta c - \Lambda(\theta)\right\}, \text{ c.à.d. solution de l'équation } \quad \frac{\partial\Lambda(\theta)}{\partial\theta} = c \\ \sigma_c^2 &= \frac{\partial^2\Lambda(\theta_c)}{\partial\theta^2} \end{aligned} \tag{5.25}$$

Dans la suite, nous présentons les deux résultats asymptotiques évoqués. Ces deux résultats sont obtenus à partir des principes de grandes déviations, ou P.G.D. Ce terme, formalisé en annexe (section A.3.3), permet d'approximer la probabilité de certains événements intéressants lorsqu'un paramètre du système est *suffisamment grand.*

5.8.2. Asymptotique du Grand Buffer

L'asymptotique de grand buffer (large buffer asymptotic) est la dénomination courante d'un principe de grandes déviations pour le travail cumulé dans une file d'attente (W), où le paramètre du système correspond à la taille ou au seuil du buffer (B). Ainsi :

$$\mathbb{P}(W > B) \approx e^{-\delta B} \tag{5.26}$$

Cet résultat revient à dire que pour des tailles de buffer considérables, le logarithme de la probabilité de déborder décroît linéairement avec la taille du buffer, avec une constante δ qui dépend des propriétés stochastiques des sources. Une analyse détaillée et des références de cette asymptotique sont données en annexe (cf. section A.6.2). Remarquons que certains modèles de trafic à mémoire longue, comme le mouvement fractionnaire Brownien, ne vérifient pas cette relation (voir p.ex. [96] ou [73]).

L'asymptotique du grand buffer est développée autour de l'idée que les débordements de buffer et/ou les pertes sont considérés comme des événements rares et donc, en général, que la taille ou le seuil des buffers est grande. Son applicabilité reste limitée quand les buffers sont dimensionnés avec des contraintes de délai. L'ingénieur réseau se pose quand même la question suivante : « à partir de quel valeur de B la taille du buffer est suffisamment grand ? ». Il est difficile de donner un réponse concrète à cette question, qui dépend de l'estimateur utilisé. Les expressions asymptotiquement correctes que l'on peut trouver dans la littérature sont $O(1/B)$.

5.8.3. Asymptotique du Grand Nombre d'usagers

Une asymptotique plus riche est *l'asymptotique du grand nombre d'usagers (ang. many sources asymptotic),* applicable quand le paramètre tendant vers l'infini (N) correspond au nombre de sources indépendantes multiplexées, la capacité du lien de transmission et la taille du buffer associé étant du même ordre de grandeur que le nombre de sources $O(N)$.

L'asymptotique du grand nombre d'usagers suppose que la taille du buffer et le nombre d'usagers multiplexés sont du même ordre de grandeur que la capacité du lien de transmission. Cette asymptotique a été développée par Weber-Courcoubetis (dans le contexte à temps discret), Botvitch-Duffield et Simonian-Guibert (dans le contexte à temps continu) sous forme d'équivalences logarithmiques. Likhanov et Mazumdar ont obtenu des expressions asymptotiquement correctes, dans un contexte à temps discret.[105][11][101][58].

Considérons un lien de transmission à conservation de travail de capacité C, modélisé par une file d'attente à buffer de taille infinie dans lequel on définit un seuil noté B. Un nombre de sources appartenant à un ensemble fini de classes $j = 1..J$ est multiplexé sur le lien de transmission. Le nombre total de source est Nn où $n = (n_1, n_2, ...n_J)$. Si nous notons $c = N^{-1}C$, $b = N^{-1}B$ et $Q(Nc, Nb, Nn) \triangleq \mathbb{P}[W \geq B]$, la probabilité que le travail cumulé dans la file d'attente dépasse le seuil B, alors :

Théorème 5.8.2 (Asymptotique de Grand nombre d'Usagers (Many Sources Asymptotic)).

$$\lim_{N\to\infty} \frac{1}{N} \log Q(Nc, Nb, Nn) = \sup_t \inf_s \left[st \sum_{j=1}^{J} n_j \alpha_j(s,t) - s(b+ct) \right] =$$
$$= -\inf_t \sup_s \left[s(b+ct) - st \sum_{j=1}^{J} n_j \alpha_j(s,t) \right] = -I$$

et donc

$$\mathbb{P}[W \geq B] \approx e^{-NI}$$

Si nous définissons un critère de performance comme une probabilité de pertes maximale $e^{-\gamma}$, noté $e^{-N\gamma_0}$, alors la relation

$$Q(Nc, Nb, Nn) \leq e^{-\gamma} = e^{-N\gamma_0}$$

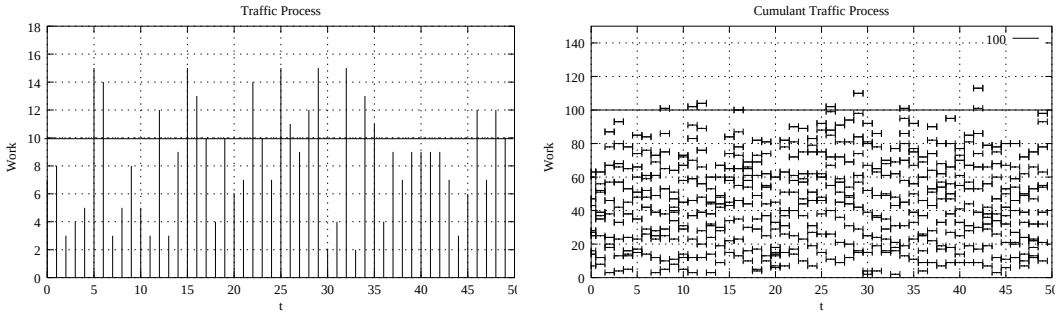
détermine un ensemble de vecteurs **n** vérifiant ce critère. Cet ensemble $A(N\gamma_0, Nc, Nb)$ est appelé région admissible, et sa caractérisation est complexe. D'après Kelly [53]

$$\lim_{N\to\infty} \frac{A(N\gamma_0, Nc, Nb)}{N} = \bigcap_{0<t<\infty} A_t \triangleq A$$
$$A_t = \left\{ (n_1, n_2, ..., n_J) : \inf_s \left[st \sum_{j=1}^{J} n_j \alpha_j(s,t) - s(b+ct) \right] \leq -\gamma_0 \right\} \tag{5.27}$$

En conséquence, pour des valeurs de N assez grandes, cette région peut être approximée par une région linéaire (affine) avec N, déterminée par :

$$\sum_{j=1}^{J} N_j \alpha_j(s,t) \leq C + \frac{1}{t}\left(B - \frac{\gamma}{s}\right) = C^* \tag{5.28}$$

Kelly étend cette région en utilisant l'amélioration de Bahadur Rao (cf. [53]).



(a) 1 Source. (b) N=10 Sources.

Fig. 5.7: Effet du multiplexage statistique et interprétation de l'asymptotique de grand nombre d'usagers. Pour des valeurs de N (facteur d'échelle : nombre de sources) *assez grandes,* les valeurs de la probabilité de pertes et du taux de pertes présentent une décroissance exponentielle avec N et la fonction de taux I. (a) Source avec une distribution uniforme [1,15], avec $c = 10$. (b) Agrégation de $N = 10$ sources i.i.d avec $C = Nc = 100$.

5.8.3.1. Principaux résultats de Likhanov et Mazumdar [58]

Comme évoqué précédemment, Likhanov et Mazumdar ont obtenu des expressions asymptotiquement exactes pour les probabilités de débordement et le taux de pertes dans certains cas particuliers. Une de leurs contributions consiste à identifier les coefficients apparaissant devant les exponentielles et permet donc, d'obtenir une amélioration notable de la précision des approximations. Pour que le lecteur puisse avoir une référence complète de l'asymptotique, nous reprenons les résultats de Likhanov et Mazumdar [58]. Le point de départ de ces résultats est le théorème de Cramér avec l'amélioration de Bahadur-Rao, et donc la caractérisation de la queue de distribution d'un agrégat de trafic.

Notation

Les points suivants définissent la notation utilisée, ainsi que les définitions des coefficients de variance et de la fonction de taux apparaissant dans le théorème 6.7.3 et les propositions 6.7.1 et 6.7.2. On note λ le débit moyen d'une source et h son débit crête. Considérons un système à temps discret $(t, n \in \mathbb{Z})$, composé d'une file d'attente de capacité C et taille de buffer B, alimentée par un agrégat de N sources i.i.d.

1. *Processus de débit instantané*

$$\begin{aligned} &x[n], \text{processus de débit instantané d'une source générique.} \\ &x^i[n], i \in 1, \ldots, N \quad \text{suivant la même loi que } x[n] \\ &x^i[n], \text{ à valeurs dans } [0, h] \end{aligned} \tag{5.29}$$

2. *Hypothèse de stabilité*

$$h > c > \lambda > 0 \tag{5.30}$$

3. *Processus d'accroissements* Le travail réalisé par la source i en $t_2 - t_1$ intervalles de temps consécutifs est noté

$$\begin{aligned} x^i(t_1, t_2] &= \sum_{m=t_1+1}^{t_2} x^i[m] \\ x^i[t_1, t_2) &= \sum_{m=t_1}^{t_2-1} x^i[m] \end{aligned} \tag{5.31}$$

4. *Processus d'accroissements d'un agrégat de trafic*

$$\begin{aligned} X^{(N)}(t_1, t_2] &= \sum_{i=1}^{N} x^i(t_1, t_2] \\ X^{(N)}[t_1, t_2) &= \sum_{i=1}^{N} x^i[t_1, t_2) \end{aligned} \tag{5.32}$$

5. *Transformée de (Log)Laplace et Fenchel-Legendre(Convexe)*

$$\begin{aligned} L_X^t(\theta) &= \mathbb{E}[e^{\theta x[0,t)}] \\ \Lambda_X^t(\theta) &= \log \mathbb{E}[e^{\theta x[0,t)}] \\ \Lambda_{X,t}^*(x) &= \sup_\theta \left\{\theta x - \Lambda_X^t(\theta)\right\} \end{aligned} \tag{5.33}$$

Notons que

$$\begin{aligned} L_X^{t'}(\theta) &= \mathbb{E}[x[0,t) e^{\theta x[0,t)}] \\ L_X^{t''}(\theta) &= \mathbb{E}[x[0,t)^2 e^{\theta x[0,t)}] \\ \mathbb{E}[X[0,t)] &= L_X^{t'}(0) \\ \mathbb{E}[X[0,t)^2] &= L_X^{t''}(0) \\ \text{var}[X[0,t)] &= L_X^{t''}(0) - \left(L_X^{t''}(0)\right)^2 \\ \frac{L_X^{t''}(\theta)}{L_X^t(\theta)} &= \frac{\partial^2 \Lambda_X^t(\theta)}{\partial \theta^2} + \left(\frac{\partial \Lambda_X^t(\theta)}{\partial \theta}\right)^2 \end{aligned} \tag{5.34}$$

6. *Coefficient de variance*

$$\sigma_t^2 = \frac{L_X^{t''}(\theta)}{L_X^t(\theta)} - (ct + b)^2 \tag{5.35}$$

7. *Fonction de taux et point de travail spatial*

$$\begin{aligned} I_t(c,b) &= \sup_{\theta} \left\{ (ct+b)\theta - \Lambda_X^t(\theta) \right\} = \\ &= (ct+b)\theta_t^* - \Lambda_X^t(\theta_t^*) = \Lambda_{X,t}^*(x) \\ \theta_t^* &= \text{argsup} I_t(c,b) \quad \text{unique solution de l'équation} \\ \frac{\partial \Lambda_X^t(\theta)}{\partial \theta} &= \frac{L_X^{t'}(\theta)}{L_X^t(\theta)} = ct+b \end{aligned} \tag{5.36}$$

Queue de Distribution Asymptotique d'un agrégat de trafic Par la suite on considère une file d'attente à temps discret. On considère un agrégat de N sources ergodiques et stationnaires i.i.d $x^i[n]$ (on note $x[n]$ la source «générique» correspondante). Le résultat suivant, concernant la Queue-Distribution Asymptotique d'un agrégat de trafic est due à Bahadur-Rao *(Bahadur-Rao improvement)* et apparaît dans [53] ou [58].

Théorème 5.8.3 (Queue de Distribution Asymptotique d'un agrégat de trafic). *Une application directe du théorème de Bahadur-Rao nous permet d'écrire :*

$$\mathbb{P}(X^{(N)}[0,t) > N(ct+b)) = \frac{1}{\theta_t^* \sqrt{2\pi\sigma_t^2 N}} e^{-NI_t(c,b)} \left(1 + O\left(\frac{1}{N}\right)\right) \tag{5.37}$$

Systèmes à buffer infini.

En appliquant l'équation de Lindley, il est possible d'obtenir des expressions asymptotiques pour la queue de distribution du travail cumulé dans une file d'attente infinie. L'équation suivante caractérise la dynamique du processus de travail cumulé et son développement récursif permet d'obtenir une équation simple pour la distribution stationnaire.

$$\begin{aligned} W_t^{(N)} &= \max\left(0, W_{t-1}^{(N)} + \sum_{i=1}^{N} x^i[t] - Nc\right) \\ W^{(N)} &= \sup_{t \in \{1,\dots\}} \left(X_{-t}^{(N)} - Nct\right) \end{aligned} \tag{5.38}$$

Proposition 5.8.1 (Queue de Distribution du travail cumulé [58]). *Sous l'hypothèse qu'il existe un t_0 vérifiant*

$$I_{t_0}(c,b) = \min_{t \in \{0,1,\dots\}} I_t(c,b) > 0$$

et

$$\liminf_{t \to \infty} \frac{I_t(c,b)}{\log t} > 0$$

alors

$$\mathbb{P}(W^{(N)} > Nb) = \frac{1}{\theta_{t_0}^* \sqrt{2\pi\sigma_{t_0}^2 N}} e^{-NI_{t_0}(c,b)} \left(1 + O\left(\frac{1}{N}\right)\right)$$

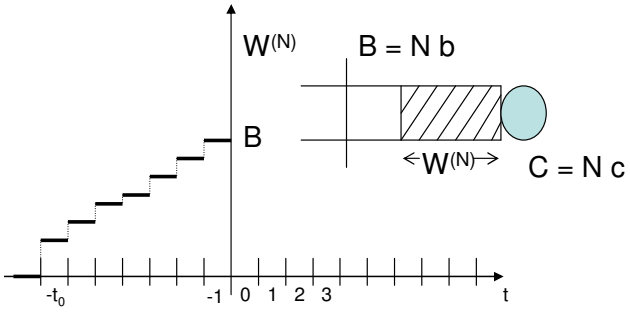


Fig. 5.8: Interprétation du paramétre t_0.

Systèmes à buffer fini

Dans des systèmes à buffer fini, l'équation de Lindley devient :

$$Y_t^{(N)} = \min\left(\max\left(0, Y_{t-1}^{(N)} + \sum_{i=1}^{N} x^i[t] - Nc\right), Nb\right) \tag{5.39}$$

Proposition 5.8.2 (Expression Asymptotique du taux de pertes [58]). *Sous l'hypothèse qu'il existe un t_0 vérifiant*

$$I_{t_0}(c,b) = \min_{t\in\{0,1,\ldots\}} I_t(c,b) > 0$$

et

$$\liminf_{t\to\infty} \frac{I_t(c,b)}{\log t} > 0$$

alors, le taux de pertes vérifie

$$\begin{aligned}\mathbb{P}_L &= \frac{\mathbb{E}\left[\max\left(0, Y_{t-1}^{(N)} + \sum_{i=1}^{N} x^i[t] - N(b+c)\right)\right]}{\mathbb{E}[\lambda_t^{(N)}]} \\ &= \frac{1}{(\theta_{t_0}^*)^2\mathbb{E}[x[i]]\sqrt{2\pi\sigma_{t_0}^2 N^3}} e^{-NI_{t_0}(c,b)}\left(1 + O\left(\frac{1}{N}\right)\right)\end{aligned}$$

Multiplexage Sans Buffer

Dans le cas où $B = 0$

$$\mathbb{P}_L = \frac{1}{(\theta_1^*)^2 c\rho\sqrt{2\pi\sigma_1^2 N^3}} e^{-NI_1(c,0)}\left(1 + O\left(\frac{1}{N}\right)\right) \tag{5.40}$$

Le multiplexage sans buffer (ou avec l'hypothèse de petits buffers) sera utilisé dans le chapitre 9,

Ingénierie de trafic.

Approximations de petits buffer

Grâce à la continuité de la fonction de taux, [58] ont obtenu des approximations du taux de pertes sous l'hypothèse de buffers à croissance sous-linéaire. Sous les hypothèses mentionnées précédemment, nous avons :

$$\mathbb{P}_L \approx \frac{1}{(\theta_1^*)^2 c\rho\sqrt{2\pi\sigma_1^2 N^3}} e^{(-NI_1(c,0)-\theta_1^* B_o(N))} \left(1 + O\left(\frac{1}{N}\right)\right) \tag{5.41}$$

Cette approximation sera utilisé dans le chapitre 6, *Partage de charge sur une topologie multi-lien*.

5.8.3.2. Remarques

- Il est intéressant de noter qu'un nombre important d'études de performance utilisent uniquement la borne de Chernov et travaillent seulement avec les équivalents logarithmiques. Comme mentionné, Likhanov et Mazumdar insistent sur l'amélioration de la précision des expressions asymptotiques comparée à l'utilisation simpliste des équivalents logarithmiques. Pour des raisons de simplicité, les résultats que nous présentons dans cette thèse utilisent ces équivalents. Néanmoins, il est facile de les compléter avec les expressions présentées ici.
- Une des hypothèses utilisées pour l'obtention des estimateurs suppose que le débit crête des sources est fini. Cependant certains modèles de trafic (citons par exemple le mouvement Brownien fractionnaire) ne vérifient pas cette hypothèse et sont néanmoins souvent utilisés [19]. Likhanov et Mazumdar ont montré que si les sources multiplexés n'ont pas un débit crête borné, la probabilité que le travail total produit par la totalité de sources de l'agrégat dépasse un seuil et des métriques dérivées comme la probabilité de débordement ou le taux de pertes ne peuvent pas présenter une décroissance exponentielle avec le nombre de sources N.
- En pratique, le débit crête des sources est intrinsèquement borné.
- Certains auteurs utilisent le terme «infsup formula», pour faire référence à la double optimisation nécessaire pour calculer la fonction de taux de l'asymptotique de grand nombre d'usagers ainsi que les équivalents logarithmiques des queues de distribution du travail cumulé et du taux de pertes.

6. Partage de charge sur une topologie multi-lien

6.1. Introduction et Motivation

Les chapitres précédents ont présenté le contexte technologique de nos travaux ainsi que les principaux outils mathématiques que nous utilisons. Les chapitres 6,7,8 et 9 présentent les principales contributions de cette thèse. Dans ce chapitre nous étudions le partage de charge entre deux éléments réseau, connectés par un certain nombre de liens (figure 6.1), que nous appelons « topologie multi-lien ». Ces liens peuvent représenter des liens physiques, indiquant que les deux éléments réseaux sont directement connectés, ou bien modéliser des LSPs ayant été établis avec une réservation stricte de ressources.

Notre premier objectif est de définir un modèle nous permettant de calculer un partage de charge optimal. L'approche suivie et les résultats développés dans la suite constituent les bases pour l'étude et l'analyse de problèmes et de topologies plus complexes, qui seront analysés dans les chapitres suivants.

Le chapitre est structuré de la façon suivante : le modèle du système est présenté dans la section 6.2. Le mécanisme proposé pour le partage de charge est donné dans la section 6.3, et la fonction objective à optimiser est présentée dans section 6.4. Dans la section 6.5, nous obtenons des conditions d'optimalité et dans la section 6.6 nous donnons quelques exemples qui illustrent notre approche pour quelques modèles de trafic. Certains résultats obtenus peuvent être généralisés par un théorème que nous donnons dans section 6.7. Finalement, dans la section 6.8 nous présentons rapidement une deuxième approche «orientée connexion» et la section 6.9 conclut le chapitre.

6.2. Modèle du Système

Dans la suite, nous appelons *trafic d'entrée* le trafic qui est offert au système dans sa globalité. Le mécanisme de partage de charge nous permettra d'obtenir une caractérisation du trafic *offert à chaque lien,* que nous appelons simplement *trafic offert.*

Trafic d'entrée

Le système est modélisé de la manière suivante (figure 6.2) : chaque lien du système est indexé par $l, l \in \{1..L\}$. Nous appelons *trafic (agrégat) d'entrée,* l'agrégation d'un nombre N de sources i.i.d. Le débit moyen d'une source sera noté λ, et son débit crête sera noté h. Afin d'être génériques par rapport au trafic d'entrée, nous utilisons la théorie des bandes passantes effectives présenté au chapitre précédent. $\alpha(s,t)$ dénote la b.p.e. d'une source faisant partie de l'agrégat d'entrée. La b.p.e. de l'agrégat est notée $\alpha_T(s,t)$, et sous les hypothèses i.i.d.$\alpha_T(s,t) = \sum_{n=1}^{N} \alpha(s,t) = N\alpha(s,t)$.

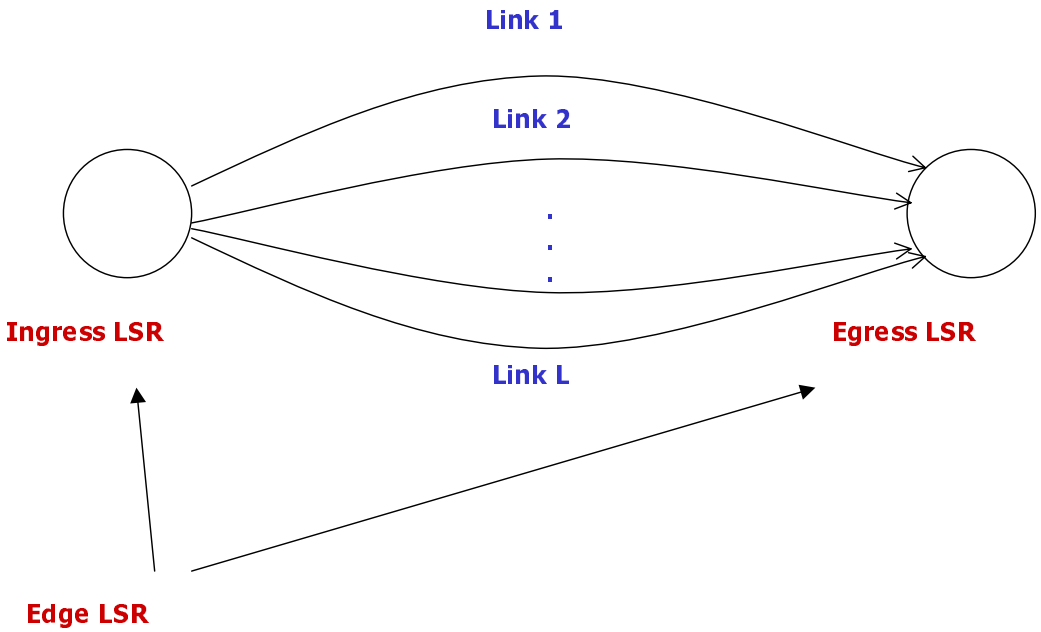


Fig. 6.1: Topologie MultiLien

Trafic offert

Nous verrons dans la section 6.3 que pour certains mécanismes de partage de charge, il est possible d'obtenir une b.p.e. *offerte* par source, qui dépendra de la b.p.e. par source d'entrée, et du mécanisme de partage de charge. Cette b.p.e. offerte au lien l sera notée $\alpha_l(s,t)$.

Trafic Externe et Contexte de Multiplexage

Nous supposons qu'à chaque lien l, le trafic offert est multiplexé avec un agrégat de trafic externe, composé de N sources i.i.d, (de distribution égale ou différente de celle du trafic d'entrée) et indépendant de celui-ci. La b.p.e. d'une source de l'agrégat externe est notée $\beta_l(s,t)$. La capacité nominale de chaque lien est notée C_l. Les tailles ou seuils des buffers sont notés B_l. La discipline est FIFO/HOL (une seule classe de trafic). Finalement, la b.p.e. totale (de l'agrégat) offerte à un lien l sous les hypothèses d'indépendance est donné par :

$$N\left(\alpha_l(s,t) + \beta_l(s,t)\right)$$

car la b.p.e. de deux agrégats indépendants est la somme de leurs b.p.e.

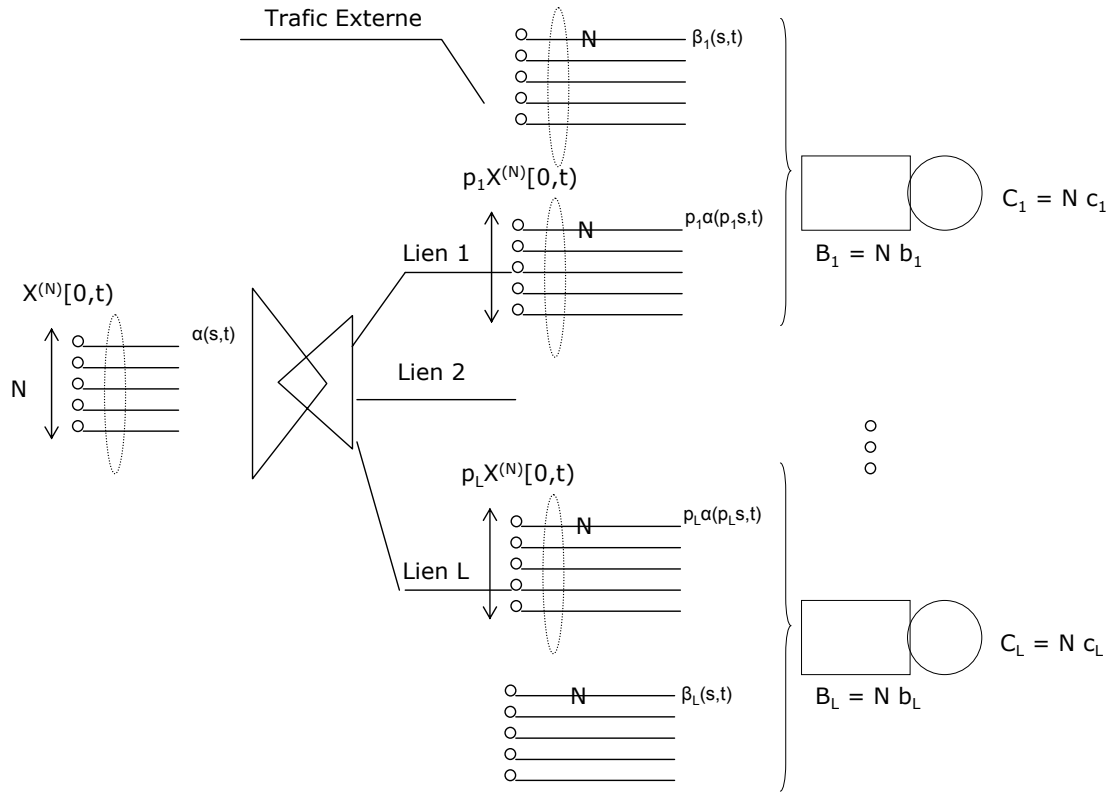


Fig. 6.2: Partage de charge sur la topologie multi-lien. Modélisation du système

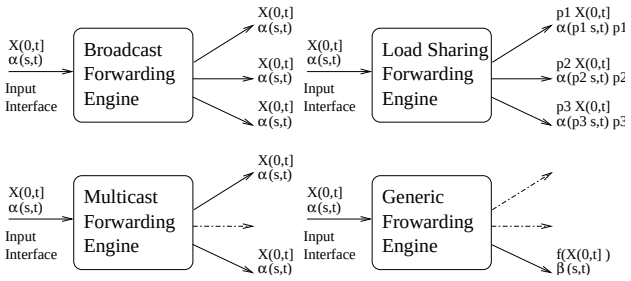
6.3. Calcul du trafic offert

6.3.1. Bandes Passantes Effectives et Acheminement parfait

Définition 6.3.1 (Mécanisme de Partage de charge). [1] Soit $\mathbb{L} = \{1, \ldots, L\}$ un groupe de LSPs et soit $X[0,t)$ un processus d'accroissements modélisant le travail produit par une source (ou un ensemble de sources). Nous définissons un mécanisme de partage de charge (ang. «load sharing scheme») comme une famille de fonctions non négatives $g_1(x), g_2(x), ..., g_L(x); g_l(x) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ vérifiant la contrainte de partage :

$$\sum_{l=1}^{L} g_l(x) = x, \qquad \forall x \geq 0$$

[1] Un résultat similaire est donné dans [13] dans le contexte de l'asymptotique du grand buffer



(a) Broadcast et Multicast.

(b) Partage de charge et acheminement générique.

Fig. 6.3: Bandes Passantes Effectives et Acheminement Parfait.

Si $g_l(X[0,t))$ est le travail offert au LSP l selon un mécanisme de partage donné (cf. figure 6.3), la b.p.e. *offerte* au LSP l est :

$$\alpha_l(s,t) = (st)^{-1} \log \mathbb{E}(e^{sg_l(X(0,t))})$$

Si $x_i[0,t)$ (resp. $X^{(N)}[0,t)$) est la quantité de travail produit par la source i (resp. la quantité de travail produit par l'agrégat de N sources) pendant l'intervalle $[0,t)$, alors $g_l(x_i[0,t))$ (resp. $g_l(X^{(N)}[0,t))$) est la quantité de travail produit par la source i (resp. quantité de travail total) *offerte* au lien l. Considérons le *mécanisme de partage linéaire,* où $g_l(x) = p_l x$. Intuitivement, p_l est la proportion déterministe de travail acheminé sur le LSP l. Remarquons que ce mécanisme correspond à une approche fluide (ce qui revient à dire que le travail est infiniment divisible).

Suivant cette approche, nous définissons par abus de langage un *partage* comme un vecteur dont la dimension dépend du nombre de liens et dont la l-ième composante représente la proportion fluide de trafic à acheminer sur le lien l. Les b.p.e. des processus de trafic par source et par agrégat offert sont données par les expressions suivantes :

$$\begin{aligned}
\alpha_l(s,t) &= \frac{1}{st} \log \mathbb{E}(e^{sp_l x[0,t)}) = p_l \alpha(p_l s, t) \\
\alpha_l^{(N)}(s,t) &= \frac{1}{st} \log \mathbb{E}(e^{sp_l X^{(N)}[0,t)}) = \\
&= \frac{1}{st} \log \mathbb{E}(e^{p_l s \sum_{n=1}^{N} x_n[0,t)}) = p_l \alpha_T(p_l s, t)
\end{aligned}$$

6.4. Critère d'optimisation et fonction objective

Après avoir détaillé notre modèle, nous allons formuler le problème du partage de charge comme un problème d'optimisation : le partage de charge optimal est le minimum sous contraintes d'un problème d'optimisation non linéaire. Les propriétés de convexité et l'applicabilité des techniques «classiques» d'optimisation vont dépendre du modèle de trafic choisi. Remarquons que si les bandes passantes effectives ont été estimées à partir de traces, il est possible que le système nécessite un

traitement entièrement numérique. D'autre part, il est possible d'obtenir pour certains modèles de trafic des formules analytiques simples et explicites permettant une étude analytique complète et détaillée.

Les propriétés simples d'optimisation des fonctions convexes a motivé le choix de ce type de fonctions comme fonctions de coût associées à des liens ou à des routes dans la littérature. Ces fonctions ont parfois été choisies en n'imposant que des contraintes de monotonie (croissance avec le trafic acheminé sur un lien) et de convexité. D'autres auteurs ont proposé des fonctions de coût dérivées du délai moyen d'une file d'attente M/M/1 [10].

Dans le contexte de nos travaux, nous proposons des fonctions de coût plus réalistes. A l'aide de la théorie des bandes passantes effectives, nous prenons en compte une famille de processus et classes de trafic, et en utilisant les principes des grandes déviations et les résultats asymptotiques donnant des bornes stochastiques à des événements de «perte» ou «débordement d'un seuil», nous proposons des fonctions de coût s'avérant mieux adaptées lorsque des garanties de QoS deviennent nécessaires. Nous définissons une fonction de coût globale (la fonction objective) notée $\mathbb{K}$, combinaison linéaire d'autant de fonctions de coût «au niveau du lien ou de la route» K_l que de liens disponibles. Les fonctions de coût de chaque route dépendent de la bande passante effective du trafic *offert* et du mécanisme de partage de charge que nous présenterons dans la suite.

Nous considérons le critère d'optimisation consistant à minimiser, pour chaque file, la probabilité de débordement d'un certain seuil et le taux de pertes. Dans le cas à buffers infinis, nous définissons un *seuil critique* pour le travail cumulé dans chaque file d'attente. Pour le cas de buffers finis, ce seuil est donné par la taille du buffer, noté B_l. Nous obtenons une approximation de la probabilité de débordement et du «taux de pertes». Il est important de noter que, même si ces deux métriques ne sont pas directement liées, [101] ont montré que l'on peut utiliser la probabilité de perte comme approximation du taux de pertes. Likhanov et Mazumdar [58] ont donné des expressions asymptotiquement exactes et ils ont prouvé qu'elles sont logarithmiquement équivalentes (à un coefficient multiplicatif près).

En utilisant l'équivalent logarithmique, nous unifions dans un même modèle la minimisation du taux de pertes et la minimisation de la probabilité de pertes : la fonction de coût par route que nous proposons donne une approximation du «taux de pertes» ou du «taux de débordement» et correspond au produit de p par l'équivalent logarithmique.

Notons que le fait de négliger ce coefficient a un effet sur la précision de l'approximation, car on surestime alors la vraie valeur du taux de pertes.

En synthèse, les équivalents logarithmiques de l'asymptotique du grand nombre d'usagers,

1. Nous donnent une approximation de la probabilité de pertes (saturation) et du taux de pertes.
2. Font intervenir les paramètres du système (capacités et buffers), et les propriétés du trafic en utilisant la b.p.e.
3. Présentent (dans certaines configurations) des propriétés de convexité par rapport aux variables de décision pouvant garantir l'optimalité globale de la solution.

Les fonctions de coût de chaque route dépendent de la bande passante effective du trafic *offert* (et donc de la bande passante effective du trafic d'entrée), de la fraction («share») offerte p_l et du trafic externe concurrent dans le lien ainsi que des capacités nominales et des tailles des buffers associés.

A la contrainte de partage il est nécessaire d'ajouter la contrainte *de stabilité* : la moyenne du trafic offert à chaque lien doit être strictement plus petite que sa capacité :

$$\lim_{s\to 0} N\left(p_l\alpha(p_l s,t)+\beta_l(s,t)\right) < C_l \quad \forall l$$

6.4.1. Formulation du problème

Le problème d'optimisation non linéaire sous contraintes peut être formalisé de la façon suivante :

$$\min_{p_1,\ldots,p_l\in[0,1]} \mathbb{K} = \min_{p_1,\ldots,p_l\in[0,1]} \sum_{l=1..L} \mathrm{K}_l(p_l)$$

où

$$\begin{aligned}
\mathrm{K}_l(p_l) &\triangleq p_l \times \exp(N\,\mathrm{J}_l(p_l)) \\
\mathrm{J}_l(p_l) &\triangleq -\inf_t \sup_s \left\{(b_l+c_l t)s - stp_l\alpha(p_l s,t) - st\beta_l(s,t)\right\} \\
&= -\inf_t I_t^l(c_l,b_l)
\end{aligned}$$

avec les contraintes :

$$\begin{aligned}
&\sum_{l=1..L} p_l = 1 \\
&0 \le p_l \le 1, \forall l \\
&\lim_{s\to 0} p_l\alpha(p_l s,t)+\beta_l(s,t) < N^{-1}C_l \quad \forall l
\end{aligned}$$

Comme certains auteurs le remarquent [18], c'est la résolution de la double optimisation (connue comme infsup formula) qui peut s'avérer compliquée. Remarquons que pour un t donné, le terme $s(b+ct) - st\alpha(s,t)$ est concave en s, car la transformée de Log Laplace est convexe. Une approche récente pour le calcul du point de travail utilisant une méthode itérative est donnée par Courcoubetis et al. [19], basée sur la substitution de trafic. *(ang. Traffic Substitution).*

6.5. Partage de charge Optimal

Une fois que nous avons formulé le problème, nous procédons à sa résolution par étapes : d'abord, nous montrons la propriété de croissance de la fonction de coût par rapport aux variables de décision p_l (malhereusement, il n'est pas en général possible de montrer la propriété de convexité). Ensuite, nous appliquons les conditions de Karush-Kuhn-Tucker pour trouver les conditions d'optimalité, en faisant l'hypothèse que les contraintes d'inégalité sont inactives (hypothèse qui sera levée ensuite). Ces conditions sont données sous forme d'un système d'équations dont la résolution nous donne le partage de charge optimal pour le critère cité dans le section précédente. Nous allons également voir que les contraintes de stabilité vont en fait limiter l'intervalle possible des valeurs de partage. Finalement, nous donnons des conditions suffisantes pour l'existence d'un optimum global

Proposition 6.5.1 (Croissance). *La fonction* $\mathrm{J}(p) : p \in [0,1] \to \mathbb{R}^+ \bigcup \{+\infty\}$ *définie par :*

$$\mathrm{J}(p) = -\inf_t \sup_s \{s(b+ct) - stp\alpha(ps,t) - st\beta(s,t)\}$$

est croissante par rapport à p.

Démonstration.

$$\begin{aligned}
\mathrm{J}(p) = \sup_t \inf_s \{stp\alpha(ps,t) + st\beta(s,t) - s(b+ct)\} \quad \text{or} \quad \forall (p_1,p_2) \in [0,1]^2, p_1 > p_2, \forall t, s > 0 \\
\log \mathbb{E}[e^{p_1 sX(0,t)}] > \log \mathbb{E}[e^{p_2 sX(0,t)}] \\
p_1 \log \mathbb{E}[e^{p_1 sX(0,t)}] > p_2 \log \mathbb{E}[e^{p_2 sX(0,t)}]
\end{aligned}$$

donc si l'on note $f(p,s,t) = \log \mathbb{E}[e^{psx[0,t)}] + st\beta(s,t) - s(b+ct)$,

$$\begin{aligned}
& f(p_1,s,t) > f(p_2,s,t) \forall s,t \\
& \Rightarrow f(p_1,s,t) > \inf_s f(p_2,s,t) \\
& \Rightarrow \inf_s f(p_1,s,t) > \inf_s f(p_2,s,t) \\
& \Rightarrow \sup_t \inf_s f(p_1,s,t) > \inf_s f(p_2,s,t) \\
& \Rightarrow \sup_t \inf_s f(p_1,s,t) > \sup_t \inf_s f(p_2,s,t)
\end{aligned}$$

□

Conditions d'Optimalité

Nous pouvons appliquer les conditions de Karush - Kuhn - Tucker :

$$\begin{aligned}
K &\triangleq \sum_{l=1}^{L} K_l(p_l) \\
K_l &\triangleq p_l \exp(NJ_l) \\
J_l &\triangleq -\inf_t \sup_s \{(b_l + c_l t)s - stp_l\alpha(p_l s,t) - st\beta_l(s,t)\}
\end{aligned}$$

Nous pouvons donner l'expression du Lagrangien étendu :

$$L(\mathbf{p},\mathbf{v},\mathbf{u},w) = \sum_{l=1}^{L} p_l \exp(NJ_l) + \sum_{l=1}^{L} u_l p_l + \sum_{l=1}^{L} v_l(1-p_l) + w\sum_{l=1}^{L}(1-p_l)$$

Supposons que les coefficients u_l et v_l associés au problème d'optimisation sont nuls (les contraintes sont inactives). Les dérivées partielles du Lagrangien s'écrivent alors :

$$\begin{aligned}
\frac{\partial L}{\partial p_l} &= e^{NJ_l(p_l)}\left(1 + p_l N \frac{\partial}{\partial p_l} J_l(p_l)\right) - w = 0 \\
\frac{\partial L}{\partial p_i} &= e^{NJ_i(p_i)}\left(1 + p_i N \frac{\partial}{\partial p_i} J_i(p_i)\right) - w = 0
\end{aligned}$$

Etant donnée l'unicité de w, les dérivées partielles de la fonction objective :

$$\frac{\partial K}{\partial p_l} = e^{NJ_l(p_l)}\left(1 + p_l N \frac{\partial}{\partial p_l} J_l(p_l)\right)$$
$$\frac{\partial K}{\partial p_i} = e^{NJ_i(p_i)}\left(1 + p_i N \frac{\partial}{\partial p_i} J_i(p_i)\right)$$

sont égales à l'optimum :

$$e^{NJ_l(p_l)}\left(1 + p_l N \frac{\partial}{\partial p_l} J_l(p_l)\right) = e^{NJ_i(p_i)}\left(1 + p_i N \frac{\partial}{\partial p_i} J_i(p_i)\right)$$

Il est alors possible d'obtenir des conditions d'optimalité locale. [10] utilisent la notion de *Longueur Egale de la Première Dérivée (ang. equal first derivative length)* :

$$\begin{cases} e^{N(Jj(p_j) - Ji(p_i))} = \frac{(1+p_i N \frac{\partial J_i(p_i)}{\partial p_i})}{(1+p_j N \frac{\partial J_j(p_j)}{\partial p_j})} & \forall i, j \in L \\ \sum_{l=1}^{L} p_l = 1 \end{cases} \tag{6.1}$$

Donc,

$$\begin{cases} J_j(p_j^*) - J_i(p_i^*) = N^{-1} \log \frac{(1+p_i^* N \frac{\partial J_i(p_i^*)}{\partial p_i})}{(1+p_j^* N \frac{\partial J_j(p_j^*)}{\partial p_j})} & \forall i, j \in L \\ \sum_{l=1}^{L} p_l^* = 1 \end{cases} \tag{6.2}$$

Validation de l'hypothèse de contraintes d'inégalité inactives

Les contraintes d'inégalité sont inactives, autrement dit, le partage optimal n'a aucune composante égale à 0 ou à 1. Intuitivement, ceci correspond à l'idée que dans notre problème il vaut toujours mieux ajouter de nouvelles files d'attente, quel que soit le seuil ou la taille de buffer associé : considérons un système avec une seule file d'attente, pour laquelle il est possible d'estimer la probabilité qu'il y ait des pertes, ou le taux de pertes. Supposons que nous ajoutions une deuxième file d'attente. Il est toujours possible de trouver une fraction, même infinitésimale de trafic (du fait de la propriété fluide de notre approche) à acheminer sur la deuxième file qui réduise le taux de pertes total du système, c'est à dire la fonction de coût. Nous allons formaliser cette idée : considérons un groupe de L liens, et prenons la probabilité de débordement comme critère. A l'optimum, la fonction de coût est :

$$\mathbb{K} = \sum_{l=1}^{L} p_l^* \mathbb{P}_{p_l^*}(W_l \geq B_l)$$

où $\mathbb{P}_{p_l^*}(W_l \geq B_l)$ est la probabilité de débordement pour le lien l pour un partage p_l. Supposons que l'on ajoute au système un nouveau lien $(L+1)$ avec pour seuil critique B_{L+1}, et supposons qu'une fraction ε du partage p_i est déviée sur le nouveau lien. Alors ,

$$\exists \varepsilon \in (0, p_i^*) \quad | \quad \mathbb{P}_{p_i^*}(W_i \geq B_i) - \mathbb{P}_{\varepsilon}(W_{L+1} \geq B_{L+1}) > 0$$

Notons que $\mathbb{P}_p(X_l > B_l)$ et $p\mathbb{P}_p(X_l > B_l)$ sont croissantes avec p,

$$p_i^*\mathbb{P}_{p_i^*}(W_i \geq B_i) - p_i^*\mathbb{P}_\varepsilon(W_{L+1} \geq B_{L+1}) > (p_i^* - \varepsilon)\mathbb{P}_{p_i^*}(W_i \geq B_i) - (p_i^* - \varepsilon)\mathbb{P}_\varepsilon(W_{L+1} \geq B_{L+1})$$

Finalement,

$$\begin{aligned} p_i^*\mathbb{P}_{p_i^*}(X_i \geq B_i) &> (p_i^* - \varepsilon)\mathbb{P}_{p_i^*}(W_i \geq B_i) + \varepsilon\mathbb{P}_\varepsilon(W_{L+1} \geq B_{L+1}) \\ &> (p_i^* - \varepsilon)\mathbb{P}_{p_i^*-\varepsilon}(W_i \geq B_i) + \varepsilon\mathbb{P}_\varepsilon(W_{L+1} \geq B_{L+1}) \end{aligned}$$

et

$$\sum_{l=1}^{L} p_l^*\mathbb{P}_{p_l^*}(W_l \geq B_l) > \sum_{\substack{l=1..L \\ l\neq i}} p_l^*\mathbb{P}_{p_l^*}(W_l \geq B_l) + (p_i^* - \varepsilon)\mathbb{P}_{p_i^*-\varepsilon}(W_i \geq B_i) + \varepsilon\mathbb{P}_\varepsilon(W_{L+1} \geq B_{L+1})$$

Remarque : En conclusion, l'ajout d'un nouveau lien diminue le coût du système. Certes, il est possible d'argumenter que le fait d'ajouter une nouvelle file représente un coût qui devrait être intégré (sous la forme d'une pénalité par exemple) dans le problème formulé. Cette extension considérerait aussi le nombre de files d'attente comme variable de décision. Etant donné que le nombre de liens (ou files) disponible est une donnée du problème, nous ne prenons pas en compte cette possibilité.

6.5.1. Conditions suffisantes de l'existence d'un minimum global sous contraintes

Proposition 6.5.2 (Condition suffisante pour l'existence d'un minimum global). *Une condition suffisante pour l'optimalité globale d'une solution de (6.1) est :*

$$N\left(\frac{\partial}{\partial p_l}J_l(p_l)\right)^2 > -\left(\frac{\partial^2}{\partial p_l^2}J_l(p_l)\right) \tag{6.3}$$

Démonstration. Afin d'étendre les conditions de minimum local à des conditions de minimum global il est suffisant de prouver la convexité de la fonction objective dans l'ensemble d'admission (l'ensemble des vecteurs de partage respectant les contraintes), et que cet ensemble est un ensemble convexe. D'après la proposition (7.5.1), le terme $\exp(NJ_l)$ est croissant avec p_l et K_l est convexe si $\exp(NJ_l)$ est convexe. Si l'on calcule la dérivée partielle seconde (par rapport à p_l) :

$$\begin{aligned} \frac{\partial^2}{\partial {p_l}^2}e^{NJ_l(p_l)} &= Ne^{NJ_l(p_l)}\left(\left(\frac{\partial^2}{\partial p_l^2}J_l(p_l)\right) + N\left(\frac{\partial}{\partial p_l}J_l(p_l)\right)^2\right) > 0 \Leftrightarrow \\ &\Leftrightarrow \left(\frac{\partial^2}{\partial p_l^2}J_l(p_l)\right) + N\left(\frac{\partial}{\partial p_l}J_l(p_l)\right)^2 > 0 \end{aligned}$$

□

Notons que l'on ne peut pas a priori garantir l'existence d'un partage optimal. Dans ce cas, l'échec de la procédure pour trouver une solution doit être signalé et éventuellement déclencher une requête pour le recalcul ou la modification du système (augmentation de la capacité, réduction du nombre de sources multiplexées, etc.)

6.6. Exemples

Dans cette section nous présentons plusieurs exemples illustrant l'approche que nous avons présentée. Les deux premiers utilisent des modèles de trafic correspondant à des sources de type Mouvement Fractionnaire Brownien (fBm) et Poisson.

Remarque importante

Nous avons vu dans le chapitre 5 que si les sources de trafic n'ont pas un débit crête borné (hypothèse non vérifiée pour ces deux types de trafic), l'asymptotique de grand nombre d'usagers ne peut pas être appliquée. Il possible de montrer que dans ce cas, le taux de pertes et la probabilité de débordement ne présentent pas une décroissance exponentielle [58]. Autrement dit, l'agrégation d'un grand nombre de sources avec débit crête non borné n'implique pas une décroissance exponentielle du travail total produit avec le nombre de sources multiplexés. Ce résultat remet en cause l'utilisation des sources de type fBm dans le contexte de l'asymptotique du grand nombre d'usagers [105].

Néanmoins, il est toujours possible d'utiliser l'expression de la bande passante effective d'une source de type fBm comme *borne supérieure* : considérons une source de trafic à débit crête borné dont la b.p.e. $\hat{\alpha}(s,t)$ vérifie :

$$\hat{\alpha}(s,t) \leq \lambda + \frac{1}{2}s\sigma^2 t^{H-1} \tag{6.4}$$

$\forall s,t \in \mathbb{R}^+$ pour des valeurs de λ, H et σ. Etant donné que la fonction de taux qui apparait dans les équivalents logarithmiques est décroissante avec $\alpha(s,t)$ nous avons donc :

$$I(.,\hat{\alpha}(s,t)) \geq I(.,\lambda + \frac{1}{2}s\sigma^2 t^{H-1}) \tag{6.5}$$

et

$$L_R(.,\hat{\alpha}(s,t)) \leq L_R(.,\lambda + \frac{1}{2}s\sigma^2 t^{H-1}) \tag{6.6}$$

Si nous considérons une file d'attente avec capacité C et taille de buffer B, alimentée par l'agrégation d'un grand nombre N de sources i.i.d, et dont la b.p.e. est bornée par la b.p.e. d'une source fBm pour certaines valeurs des paramètres λ, σ et H, alors le système présente un taux de pertes qui est borné par le taux de pertes qu'on approxime en utilisant l'expression de la b.p.e. d'un processus fBm : $I(.,\hat{\alpha}(s,t)) \geq I(.,\lambda + \frac{1}{2}s\sigma^2 t^{H-1})$.

En synthèse, l'utilisation des modèles de trafic fBm nous permettra l'analyse du comportement du système et de son optimum à l'aide d'un nombre reduit de paramètres.

Finalement, le troisième exemple que nous présentons correspond à un cas plus réaliste, d'application immédiate, où les sources sont caractérisées par leurs débits moyens et leurs débits crêtes, et où nous utilisons une borne supérieure de la b.p.e.

Remarque

Si les sources constituant l'agrégat sont à débit constant, tout partage respectant les contraintes de stabilité (débit moyen offert à n'importe quel lien inférieur à la capacité nominale du lien) est optimal : le travail cumulé dans les différentes files d'attente est toujours zéro et les pertes sont

inexistantes. Pourtant, et afin d'être en accord avec les résultats que nous allons présenter par la suite, certaines règles de dimensionnement pourront quand même être appliquées.

6.6.1. Exemple I : Partage De Charge Multi Chemin à un seul saut (Single Hop Multipath Load Sharing) avec des sources mouvement fractionnaire Brownien

Résultats préliminaires

La suite $\{z(t), t \geq 1\}$ avec $Z(t) = \sum_{s=1}^{t} z(s)$ est un *bruit fractionnaire Gaussien normalisé de paramètre de Hurst H*[73] (Norros) si :

1. $\{z(t), t \geq 1\}$ est stationnaire.
2. $\{z(t), t \geq 1\}$ est une suite Gaussienne.
3. $\mathbb{E}[Z(t)] = 0$
4. $\mathbb{E}[Z(t)^2] = t^{2H}$
5. Pour $t_1 < t_2 < t_3 < t_4$, l'autocovariance

$$\begin{aligned} &Cov[Z(t_2) - Z(t_1), Z(t_4) - Z(t_3)] \\ &= \mathbb{E}[(Z(t_2) - Z(t_1))(Z(t_4) - Z(t_3))] \\ &= \frac{1}{2}\left((t_4 - t_1)^{2H} - (t_3 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H}\right) \end{aligned}$$

Son autocovariance est donnée par :

$$\begin{aligned} r(n) &= Cov[Z(1), Z(n+1) - Z(n)] \\ &= \frac{1}{2}\left((n+1)^{2H} - (n)^{2H} + (n-1)^{2H} - (n)^{2H}\right) \\ &\approx H(2H-1)n^{-2(1-H)} \end{aligned}$$

Donc, pour $1/2 < H < 1$ le processus est à mémoire longue.

[73] et plus tard Duffield et O'Connell [25] considéraient le processus d'accroissements

$$X[0,t) = mt + \sqrt{ma}Z(t), \quad \text{avec } m > 0, \text{ et } a > 0$$

et ont montré que

$$\lim_{x\to\infty} \frac{1}{x^{2(1-H)}} \log \mathbb{P}\left(W \geq x\right) = -\frac{1}{2am(1-H)^2}\left((c-m)\frac{1-H}{H}\right)^{2H} \tag{6.7}$$

Remarquons que pour $H > 1/2$, l'applicabilité de l'asymptotique du grand buffer est limitée, car la queue de distribution n'est pas exponentielle (le logarithme de la probabilité ne décroît pas linéairement avec la taille du buffer). D'autre part, dans son article sur les b.p.e., Kelly [53] considère également des processus d'accroissements $Y(0,t) = \lambda t + V(t)$ où $V(t)$ est une variable aléatoire normale de moyenne zéro, et il insiste sur le fait que la forme analytiquement simple des b.p.e. des sources gaussiennes et l'obtention de formules simples, justifie le problème des accroissements négatifs. Le cas d'une source modélisée comme un mouvement fractionnaire Brownien *(ang. Fractional Brownian*

Motion ou fBm) correspond au cas particulier où la variance de $V(t)$ est $\sigma^2 t^{2H}$, où σ et H sont des constantes (notons que $\sigma^2 = ma$). L'auteur étudie la superposition d'un grand nombre de sources fBm appartenant à J classes différentes : $\mathbf{N} = N(n_1, \ldots, n_J)$. Pour une probabilité de débordement limite de $e^{-N\gamma}$, la région d'admission *(ang. acceptance region)* est définie par l'équation :

$$H\left(\frac{1-H}{b}\right)^{\frac{1}{H}-1}\left(2\gamma\sum_{j\in J} n_j\sigma_j^2\right)^{\frac{1}{2H}} + \sum_{j\in J} n_j\lambda_j \leq c$$

et le point de travail :

$$s^* = 2\,(1-H)\,\frac{\gamma}{b}$$
$$t^* = \left(\frac{H}{1-H}\right)\frac{b}{c-\sum\limits_j n_j^*\lambda_j}$$

Appliquons ce modèle de trafic à notre approche. Considérons le trafic d'entrée comme l'agrégation d'un nombre N de sources fBm, indépendantes, de même paramètre de Hurst H. La bande passante effective de chaque source et de l'agrégat sont données par les expressions suivantes :

$$\alpha_m(s,t) = \lambda + \frac{1}{2}\sigma^2 s t^{2H-1}$$
$$\alpha_T(s,t) = \sum_{m=1}^{N}\lambda + \frac{1}{2}\sum_{m=1}^{N}\sigma^2 s t^{2H-1} = N\lambda + \frac{1}{2}N\sigma^2 s t^{2H-1}$$

Remarquons que l'agrégation de sources fBm (i.i.d) est une source fBm de moyenne $N\lambda$ et de variance $N\sigma^2$. Sur chaque lien l de la topologie, le trafic étudié est multiplexé avec un agrégat de trafic externe, composé d'un nombre N de sources fBm, de même paramètre de Hurst H. La b.p.e. d'une source faisant partie du trafic externe de la file l, notée $\beta_l(s,t)$, est donné par :

$$\beta_l(s,t) = \gamma_l + \frac{1}{2}\tau_l^2 s t^{2H-1}$$

La b.p.e. associée à l'agrégat est

$$\beta_T(s,t) = \sum_{m=1}^{N}\gamma_l + \frac{1}{2}\sum_{m=1}^{M}\tau_l^2 s t^{2H-1} = N\gamma_l + \frac{1}{2}N\tau_l^2 s t^{2H-1}$$

Remarquons que d'après [19], n'importe quel trafic externe (pour le calcul de la fonction de taux) peut être remplacé par une source fBm, même s'il n'est pas à mémoire longue.

La fonction de taux par file correspond à :

$$\begin{aligned} I_l &= \inf_t \sup_s \{(b_l + c_l t)s - st\lambda p_l - st\frac{1}{2}\sigma^2 p_l^2 s t^{2H-1} - st\gamma_l - st\frac{1}{2}\tau_l^2 s t^{2H-1}\} \\ &= \inf_t \sup_s \{b_l s + (c_l - \lambda p_l - \gamma_l)st - \frac{1}{2}(\sigma^2 p_l^2 + \tau_l^2)s^2 t^{2H}\} \end{aligned}$$

Les différents points de travail pour chaque lien sont donnés par :

$$s_l^* = \frac{b_l + (c_l - \gamma_l - \lambda p_l)t}{(\sigma^2 p_l^2 + \tau_l^2)t^{2H}}$$
$$t_l^* = \frac{b_l}{c_l - \gamma_l - \lambda p_l}\frac{H}{1-H}$$

Ce qui nous donne

$$J_l = -\frac{1}{2(\sigma^2 p_l^2 + \tau^2)}\frac{b_l^{2(1-H)}(\frac{1}{1-H})^{2(1-H)}}{H^{2H}}(c_l - \gamma_l - \lambda p_l)^{2H}$$

Considérons, pour alléger l'exposé, $H = 1/2$:

$$J_l = -\frac{2b_l(c_l - \gamma_l - \lambda p_l)}{\sigma^2 p_l^2 + \tau^2} \tag{6.8}$$

Les conditions d'optimalité deviennent :

$$\begin{cases} e^{-2N\frac{b_i(c_i-\gamma_i-\lambda p_i)}{\sigma^2 p_i^2+\tau_i^2}}\left(1 - p_i 2N b_i \frac{\left(\lambda(\sigma^2 p_i^2 - \tau_i^2) - 2\sigma^2 p_i(c_i-\gamma_i)\right)}{(\sigma^2 p_i^2+\tau_i^2)^2}\right) = \\ e^{-2N\frac{b_j(c_j-\gamma_j-\lambda p_j)}{\sigma^2 p_j^2+\tau_j^2}}\left(1 - p_j 2N b_j \frac{\left(\lambda(\sigma^2 p_j^2 - \tau_j^2) - 2\sigma^2 p_j(c_j-\gamma_j)\right)}{(\sigma^2 p_j^2+\tau_j^2)^2}\right) \\ \forall i, j \in L \\ \sum_{l=1}^{L} p_l = 1 \end{cases}$$

Remarque

Remarquons l'analogie entre l'expression 6.8 et le résultat de Norros (eq. 6.7), où pour $H \approx 1/2$ correspond à l'asymptotique du grand buffer.

6.6.1.1. Analyse

Fonction de Coût par Lien

Afin d'illustrer l'effet des différents paramètres sur les fonctions de coût proposées, la figure 6.4 montre l'évolution de la fonction de coût par lien (partage multiplié par l'équivalent logarithmique) avec des sources mouvement fractionnaire Brownien en considérant $\gamma = 0$ et $\tau = 0$ (système isolé, sans trafic externe) en fonction du partage p alloué au lien, pour différentes tailles de buffer B, avec $C = 300, \lambda = 9.9, H = 0.5$ et $\sigma = 3$. En accord avec l'intuition, des liens avec une taille de buffer plus grande (plus permissifs) induisent un coût inférieur. Remarquons qu'avec la configuration choisie c'est lorsque p tend vers 1 que le débit moyen par source offert est proche de $c = N^{-1}C$. La figure 6.5(a) montre l'effet de la variance (σ^2) et figure 6.5(b) montre l'effet du paramètre de Hurst. Ainsi, on peut noter comment la sporadicité des sources est prise en compte : un agrégat avec des valeurs de σ^2 décroissantes se comporte de plus en plus comme des sources à débit instantané constant et, si l'on respecte les contraintes de stabilité, les fonctions de coût sont de plus en plus plates (sauf à proximité de la charge critique) et les probabilités de perte ou le taux de pertes tendent vers zéro.

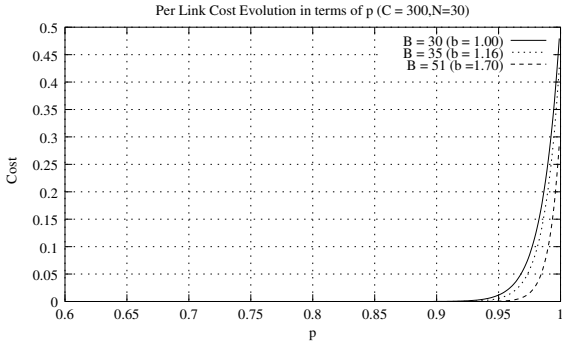
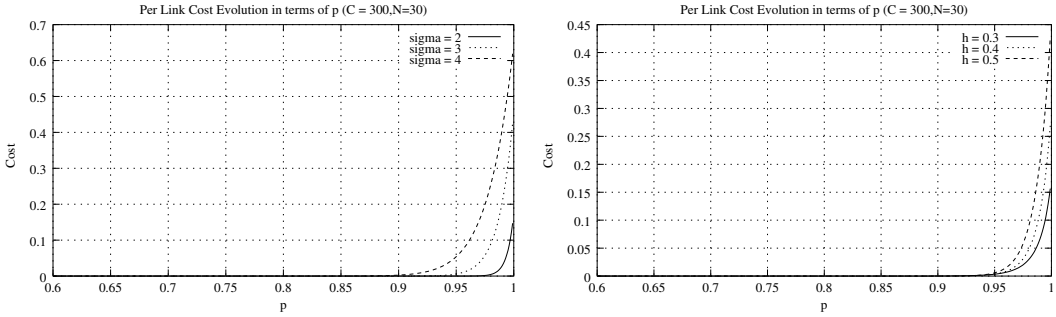


Fig. 6.4: Sources fBm : Fonction de coût par lien, effet de B



(a) Effet de σ (b) Effet de H

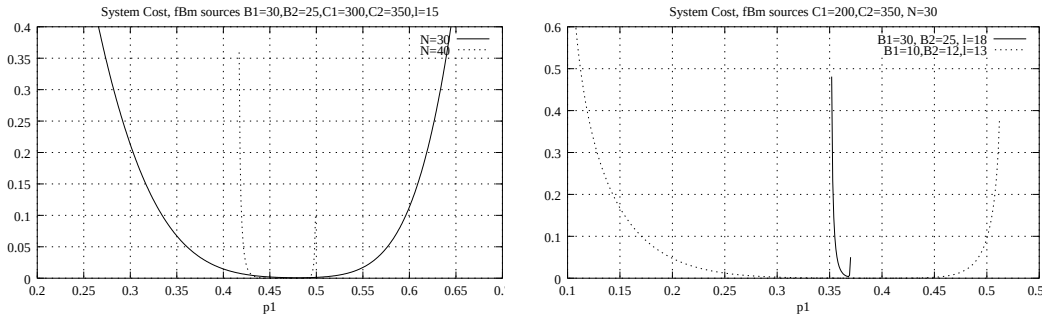
Fig. 6.5: Sources fBm : Fonction de coût par lien, effet de H et σ

Fonction de Coût du Système

Une fois caractérisées les fonctions de coût par lien, nous donnons deux exemples de la fonction de coût globale, pour un système à deux liens ($L = 2$). Dans la première configuration (figure 6.6 (a)), $C_1 = 300, B_1 = 30, C_2 = 355, B_2 = 25, \lambda = 15, H = 0.5$ pour $N = 30, 40$, et l'optimum est pour $N = 30$,$(0.48, 0.52)$ et pour $N = 40$,$(0.473, 0.527)$. Dans la deuxième configuration, (figure 6.6 (b)), $C_1 = 200, B_1 = 30, C_2 = 355, B_2 = 25, H = 0.5, N = 30$ pour $\lambda = 13, 18$, et l'optimum est pour $\lambda = 18$,$(0.368, 0.632)$ et pour $\lambda = 13$,$(0.397, 0.603)$. Remarquons que dans certaines configurations, les contraintes de stabilité peuvent fortement limiter l'ensemble des possibles.

6.6.1.2. Approximations et «règles simples»

En rajoutant certaines hypothèses, il est possible de simplifier les conditions d'optimalité et d'en déduire des règles intuitives simples. Un premier exemple trivial correspond au cas où $\tau_l^2 \ll \gamma_l \forall l = 1..L$; autrement dit, le trafic externe se comporte comme des sources à debit constant. Ce cas se rammène à un cas isolé, sans trafic externe, où les *capacités résiduelles* $r_l = c_l - \gamma_l$ jouent le rôle des



(a) $C_1 = 300, B_1 = 30, C_2 = 350, B_2 = 25, \lambda = 15, H = 0.5$ pour $N = 30, 40$

(b) $C_1 = 200, B_1 = 30, C_2 = 350, B_2 = 25, H = 0.5, N = 30$ pour $\lambda = 13, 18,$

Fig. 6.6: Sources fBm : Fonctions de coût du système

capacités nominales.

Approximation à faible charge

Une approximation grossière est $\forall l, r_l \gg \lambda$. Dans ce cas, le partage optimal correspond à l'expression suivante :

$$p_i^* = \frac{\sqrt{b_i r_i}}{\sum_{l=1}^{L} \sqrt{b_l r_l}} \tag{6.9}$$

Il est clair que l'hypothèse $\forall i, R_i \gg \lambda$ n'est pas toujours vérifiée. Nous allons voir que le partage optimal pourra néanmoins être approximé par l'expression 6.9. D'autre part, sous l'hypothèse $\tau_l^2 \ll \gamma_l \quad \forall l = 1..L$,si les tailles des buffers b_l sont proportionnelles aux capacités résiduelles, $b_l = dr_l \quad \forall l$, (d correspond à un délai virtuel maximal), alors les conditions d'optimalité deviennent :

$$e^{-2N\frac{dr_i(r_i-\lambda p_i)}{\sigma^2 p_i^2}}\left(1 + 2Ndr_i\frac{(2r_i - \lambda p_i)}{\sigma^2 p_i^2}\right) =$$
$$e^{-2N\frac{dr_j(r_j-\lambda p_j)}{\sigma^2 p_j^2}}\left(1 + 2Ndr_j\frac{(2r_j - \lambda p_j)}{\sigma^2 p_j^2}\right) \quad \forall i, j \in L$$

Autrement dit, le «délai virtuel» étant fixé, une augmentation de la capacité implique une augmentation de la taille du buffer. Dans ce cas, le partage de charge optimal correspond au cas pragmatique et intuitif de *Proportionnel à la capacité résiduelle.* D'un point de vue mathématique, ce cas correspond à celui où l'égalité est vérifiée au même temps par l'exponentielle et par le terme entre parenthèses. Ainsi :

$$p_i^* = \frac{c_i - \gamma_i}{\sum_{l=1}^{L} (c_l - \gamma_l)}$$

Nous reviendrons sur ce résultat dans la section 6.7.

6.6.2. Exemple II : Partage De Charge Multi Chemin à un seul saut (Single Hop Multipath Load Sharing) avec des sources Poissoniennes

Dans ce deuxième exemple, nous supposons qu'il n'existe pas de trafic externe et que l'agrégat de trafic d'entrée correspond à la superposition d'un nombre important de sources indépendantes modélisées par des processus de Poisson. Rappelons que la b.p.e. d'un processus d'arrivées de Poisson de paramètre λ est donnée par l'expression suivante :

$$\alpha(s,t) = \lambda\left(\frac{e^s - 1}{s}\right)$$

Ainsi, la b.p.e. de l'agrégat d'entrée, $\alpha_T(s,t)$, vérifie :

$$\alpha(s,t) = \sum_{m=1}^{N} \lambda\left(\frac{e^s - 1}{s}\right) = N\lambda\left(\frac{e^s - 1}{s}\right)$$

Analysons les fonctions de taux de chaque file d'attente. Si nous définissons

$$H_l \triangleq (b_l + c_l t)s - \lambda t(e^{p_l s} - 1)$$

le point de travail s_l^* est solution de l'equation $\partial H_l/\partial s = 0$, donc :

$$s_l^* = \frac{1}{p_l}\log\left(\frac{b_l + c_l t}{t_l \lambda p_l}\right)$$

En substituant, nous obtenons

$$I_t^l(c_l, b_l) = \sup_s H_l = \frac{1}{p_l}\left(\log\left(\frac{b_l + c_l t}{t_l \lambda p_l}\right)(b_l + c_l t) - (b_l + (c_l - \lambda p_l)t)\right)$$

Proposition 6.6.1 (Convexité). *La fonction $I_t^l(c_l, b_l)$ est convexe par rapport à $t > 0$.*

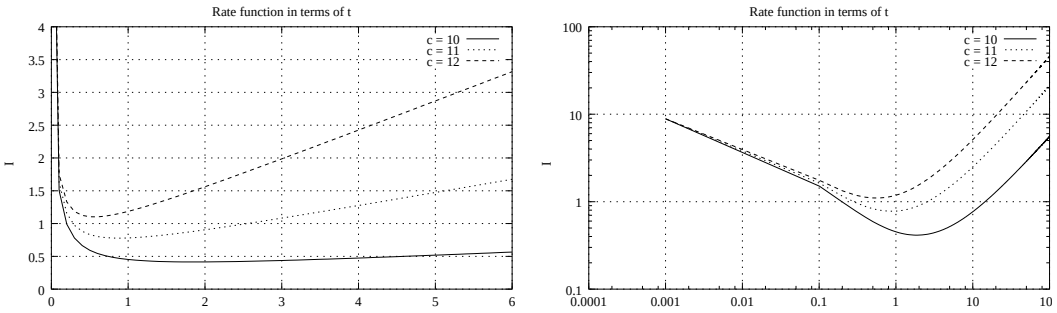
Démonstration.

$$\begin{aligned} \frac{\partial I_t^l(c_l,b_l)}{\partial t} &= \frac{1}{tp_l}\left(c_l t\log\left(\frac{b_l + c_l t}{t_l\lambda p_l}\right) - (b_l + (c_l - \lambda p_l)t)\right) \\ \frac{\partial^2 I_t^l(c_l,b_l))}{\partial t^2} &= \frac{b_l^2}{(b_l + c_l t)t^2 p_l} > 0, \quad \text{pour } t > 0, p > 0, c > 0, b > 0 \end{aligned} \tag{6.10}$$

□

Remarquons que

$$\lim_{t\to\infty}\frac{I_t^l(c_l,b_l)}{t} = \frac{1}{p_l}\left(c_l\log\left(\frac{c_l}{\lambda p_l}\right) - (c_l - \lambda p_l)\right)$$



(a) Echelle linéaire.

(b) Echelle logarithmique.

Fig. 6.7: Fonction de taux $I_t^l(c_l, b_l)$ en fonction de t $(p = 1, b = 2, \lambda = 9)$

La figure 6.7 illustre l'évolution en fonction de t pour quelques configurations. Notons le caractère linéaire quand $t \to \infty$. L'argmin de la fonction de taux correspond au point de travail $t^* > 0$ et est solution de :

$$\log\left(\frac{b_l + c_l t}{t_l \lambda p_l}\right) c_l t = (b_l + (c_l - \lambda p_l)t) \tag{6.11}$$

Finalement, la fonction de taux peut s'exprimer comme :

$$\inf_t I_t^l(c_l, b_l) = \frac{b_l}{p_l} \log\left(\frac{b_l + c_l t_l^*}{t_l^* \lambda p_l}\right) \tag{6.12}$$

et la fonction objective comme :

$$\begin{aligned}
\mathbb{K} &= \sum_{l=1}^{L} p_l \times e^{-N \frac{b_l}{p_l} \log\left(\frac{b_l + c_l t_l^*}{t_l^* \lambda p_l}\right)} \\
&= \sum_{l=1}^{L} p_l \times e^{-\frac{B_l}{p_l} \log\left(\frac{b_l + c_l t_l^*}{t_l^* \lambda p_l}\right)} \\
&= \sum_{l=1}^{L} p_l \times \left(\frac{b_l + c_l t_l^*}{t_l^* \lambda p_l}\right)^{-\frac{B_l}{p_l}}
\end{aligned}$$

où le point de travail temporel t^* est donné par l'équation (6.11).

6.6.2.1. Analyse

Fonction de Coût par Lien

La figure 6.9 montre l'évolution de la fonction de coût par lien (partage multiplié par l'équivalent logarithmique) en fonction de la proportion de trafic allouée au lien, p pour différentes tailles de buffer B, avec $N = 30$ (a) $\lambda = 9.9$ et (b) $\lambda = 14$ unités de travail par unité de temps.

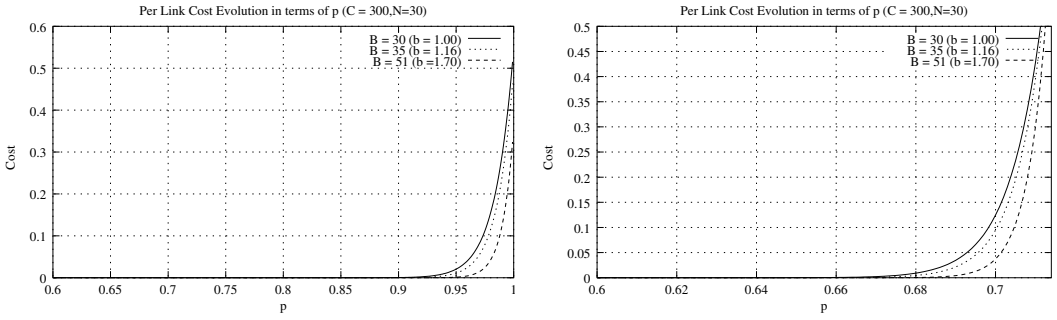
```
001 namespace effbw { namespace poisson {
002
003 ////////////////////////////////////////////////////////////////
004 // Calcul du numérateur de la dérivée de la fonction de taux
005 // p.r.à t, sources Poisson.
006 ////////////////////////////////////////////////////////////////
007 template<typename real_t> inline
007 real_t RateDerivativeT (real_t t, real_t b, real_t c, real_t lambda, real_t p)
008 {
009   return ( c * t * log (  (b + c * t) / (lambda * t *p))  - ( b +  (c - lambda * p) *
t)) ;
010 }
011
012 ////////////////////////////////////////////////////////////////
013 // Calcul du point de travail t*, sources Poisson
014 // par dychotomie dans [ta,tb]
015 ////////////////////////////////////////////////////////////////
016 template<typename real_t>
017 real_t OptimalT (real_t b, real_t c, real_t lambda, real_t p)
018 {
019   real_t ta = EPSILON ;
020   real_t tb = MAXTIME ;
021   real_t derivative_a  = RateDerivativeT (ta, b, c, lambda,  p) ;
022   real_t derivative_b  = RateDerivativeT (tb, b, c, lambda,  p) ;
023   assert (derivative_a <= 0.0) ;
024   assert (derivative_b >= 0.0) ;
025   do
025   {
026     real_t t = 1.0 / 2.0 * (ta + tb ) ;
027     real_t der  = RateDerivativeT (t, b, c, lambda,  p) ;
028     if (der > 0)
029       tb = t ;
030     else
030       ta = t ;
031   }
032   while (fabs (ta - tb) > EPSILON) ;
033   return (tb) ;
034 }
035 }}
```

Fig. 6.8: Calcul Numérique du point de travail t^*

Effet de la taille des buffers

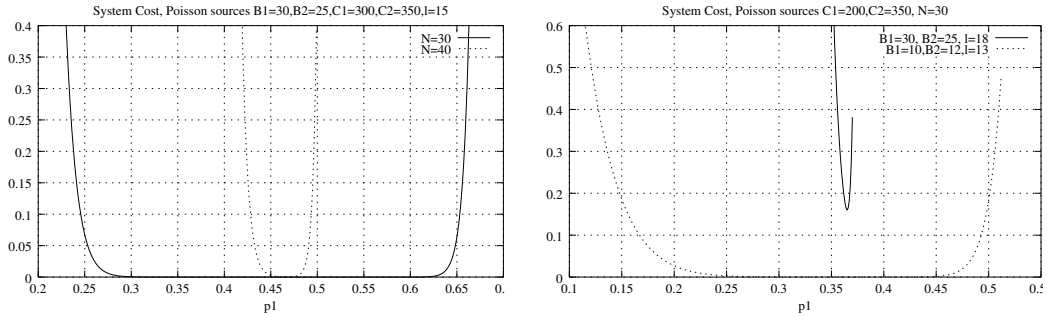
Nous reprenons la même configuration que dans le cas avec des sources de type fBm, avec des sources Poissoniennes (cf. figure 6.10).

La figure 6.11 (a) illustre l'évolution du partage de charge optimal pour un système avec $L = 2$, pour $N = 40, B_1 = 20, C_1 = C_2 = 350$ et $\lambda = 15$. Lorsque $B_2/B_1 = 1$ le système est symétrique et correspond à un partage optimal $(0.5, 0.5)$. En augmentant la taille du buffer du deuxième lien, le partage optimal favorise celui-ci. La figure 6.11 (b) illustre l'évolution du partage de charge optimal pour le même système, mais avec $\lambda = 10$. Remarquons que la variation du point optimal est beaucoup plus notable à faible charge, comparée à celle que nous pouvons évaluer à plus forte charge. A priori, il paraît raisonnable de dire qu'à forte charge, l'effet des différentes tailles des buffers devrait faire



(a) $\lambda = 9.9$ (b) $\lambda = 14$

Fig. 6.9: Trafic Poissonien : Fonction de coût par lien



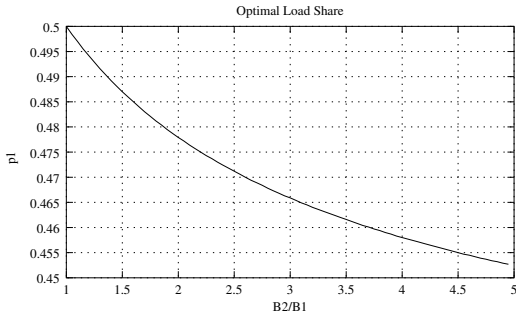
(a) (b)

Fig. 6.10: Trafic Poissonien : Fonctions de coût du système

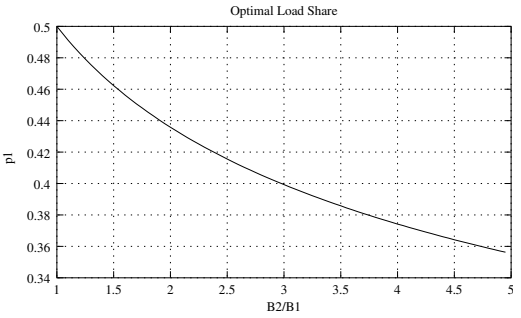
changer le partage optimal. L'explication en est simple : à forte charge, les contraintes de stabilité limitent fortement l'ensemble réalisable. Nous pouvons affirmer, encore une fois, qu'à forte charge les contraintes de stabilité vont fortement influencer le partage optimal.

Effet de la capacité

Nous illustrons sur la figure 6.12 (a) la variation du partage optimal dans le cas d'un sytème avec $L = 2, N = 30, B_1 = B_2 = 20, C_1 = 350$ et $\lambda = 21$ en fonction de la proportion C_2/C_1, à tailles de buffers constantes, ainsi que les approximations basées sur la règle de la racine carrée et sur la proportionnalité. La figure 6.12 (b) montre l'evolution du coût optimal pour ces trois cas.

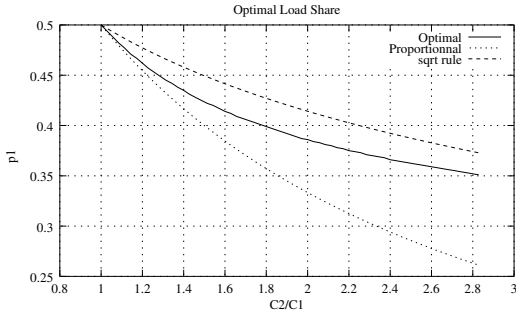


(a) $L = 2, N = 40, B_1 = 20, C_1 = C_2 = 350, \lambda = 15, B_2/B_1 = 1$

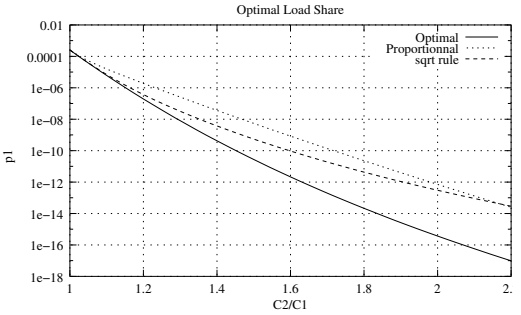


(b) $L = 2, N = 40, B_1 = 20, C_1 = C_2 = 350, \lambda = 10, B_2/B_1 = 1$

Fig. 6.11: Trafic Poissonien : Evolution du partage optimal en fonction de la taille des buffers



(a) Evolution de l'optimum, $L = 2, N = 30, B_1 = B_2 = 20, C_1 = 350$ et $\lambda = 21$



(b) Evolution de la fonction objective, $L = 2, N = 30, B_1 = B_2 = 20, C_1 = 350$ et $\lambda = 21$

Fig. 6.12: Trafic Poissonien : Evolution du partage optimal en fonction du rapport de capacités.

6.6.3. Exemple III : Partage De Charge Multi Chemin à un seul saut (Single Hop Multipath Load Sharing) avec des sources «de Hoeffding»

Dans ce troisième exemple, nous considérons des processus de trafic à temps discret, en reprenant les notations du chapitre 5. Nous allons considérer un agrégat de sources dont on connait le débit moyen (noté m ou λ) et le débit crête (noté h). Les propriétés de croissance et de convexité de la fonction exponentielle nous permettent, en utilisant l'inégalité de Hoeffding, de proposer une borne supérieure de la b.p.e. des sources.

Pour une fonction convexe et croissante f, et pour une v.a. $\Lambda = \frac{X(0,t)}{t} \in (x_{\min}, x_{\max})$ nous avons (inégalité de Hoeffding) :

$$f(\Lambda) \leq \frac{x_{\max} - \Lambda}{x_{\max} - x_{\min}} f(x_{\min}) + \frac{\Lambda - x_{\min}}{x_{\max} - x_{\min}} f(x_{\max})$$

Si le débit instantané est borné par $x_{\min}$ et $x_{\max}$ nous avons

$$\begin{aligned}
e^{sX(0,t)} &\leq \frac{x_{\max}t - X(0,t)}{x_{\max}t - x_{\min}t} e^{sx_{\min}t} + \frac{X(0,t) - x_{\min}t}{x_{\max}t - x_{\min}t} e^{sx_{\max}t} = \\
\mathbb{E}(e^{sX(0,t)}) &\leq \frac{1}{x_{\max} - x_{\min}} \left(e^{sx_{\min}t} (x_{\max} - \lambda) - e^{sx_{\max}t} (x_{\min} - \lambda) \right) = \\
&= \frac{x_{\max} - \lambda}{x_{\max} - x_{\min}} e^{sx_{\min}t} + \frac{\lambda - x_{\min}}{x_{\max} - x_{\min}} e^{sx_{\max}t} \\
\alpha(s,t) &\leq \frac{1}{st} \log \left(\frac{x_{\max} - \lambda}{x_{\max} - x_{\min}} e^{sx_{\min}t} + \frac{\lambda - x_{\min}}{x_{\max} - x_{\min}} e^{sx_{\max}t} \right)
\end{aligned}$$

Un cas très utilisé est le cas où $x_{\min} = 0$ et $x_{\max} = h$ (le débit crête). Cette borne supérieure apparaît par exemple dans [53] ou [18]. Rappelons que cette borne s'applique s'il est possible de définir un débit crête instantané fini (mathématiquement, si le supremum essentiel de $X(0,t)/t$ est fini).

$$\hat{\alpha}(s,t) = \frac{1}{st} \log \left(1 - \frac{\lambda}{h} + \frac{\lambda}{h} e^{sht} \right)$$

Utilisant notre caractérisation de la b.p.e. offerte à chaque lien, nous pouvons écrire

$$\hat{\alpha}_l(s,t) = p_l \frac{1}{p_l st} \log \left(1 - \frac{\lambda}{h} + \frac{\lambda}{h} e^{p_l sht} \right)$$

Lemme 6.6.1 (Convexité de la borne supérieure de la b.p.e.). *La b.p.e. offerte à chaque lien est convexe par rapport à p.*

Démonstration.

$$\begin{aligned}
\hat{\alpha}_l(s,t)|_{m,h} &= p_l \frac{1}{p_l st} \log \left(1 + \frac{\lambda}{h} \left(e^{p_l sht} - 1 \right) \right) \\
&= \frac{1}{st} \log \left(1 + \frac{m}{h} \left(e^{p_l sht} - 1 \right) \right) \\
&= \kappa \log \left(1 + \beta \left(e^{p\gamma} - 1 \right) \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \alpha_l(p)}{\partial p} &= \kappa \frac{\beta\gamma e^{p\gamma}}{1 + \beta \left(e^{p\gamma} - 1 \right)} \\
\frac{\partial^2 \alpha_l(p)}{\partial p^2} &= \kappa \frac{\beta\gamma^2 e^{p\gamma} (1 + \beta \left(e^{p\gamma} - 1 \right)) - \beta\gamma e^{p\gamma} \left(\beta\gamma e^{p\gamma} \right)}{\left(1 + \beta \left(e^{p\gamma} - 1 \right) \right)^2}
\end{aligned}$$

$$\begin{aligned}
\Delta &= \beta\gamma^2 e^{p\gamma} + \beta^2\gamma^2 e^{2p\gamma} - \beta^2\gamma^2 e^{p\gamma} - \beta^2\gamma^2 e^{2p\gamma} \\
&= \beta\gamma^2 e^{p\gamma} (1 - \beta) \\
&> 0 \iff \beta < 1 \quad \text{par définition, car } \beta = \frac{m}{h} < 1
\end{aligned}$$

□

Les tailles des buffers de chaque lien ne croissent pas linéairement avec le nombre de sources multiplexées, mais avec sa racine carré. Ainsi, nous écrivons :

$$\begin{aligned}
C_l &= N c_l \\
B_l &= \sqrt{N} b_l
\end{aligned}$$

D'après Likhanov et Mazumdar, dans les cas où les buffers ont une croissance sous linéaire, nous pouvons utiliser l'approximation des petits buffers [58] (cf. chapitre 5, *Outils Mathématiques*). Pour une file d'attente générique avec une capacité par source c et un buffer de taille $B = \sqrt{N}b$ nous avons :

$$\mathbb{P}_L \approx \frac{1}{(\theta_1^*)^2 c\rho\sqrt{2\pi\sigma_1^2 N^3}} e^{(-NI_1(c,0)-\theta_1^* B_o(N))}\left(1+O\left(\frac{1}{N}\right)\right) \tag{6.13}$$

où

$$I_t(c,b) = \sup_s \left\{(b+ct)s - \log\left(1-\frac{m}{h}+\frac{m}{h}e^{spht}\right)\right\}$$
$$B_o(N) = \sqrt{N}b$$

Dans les problèmes d'optimisation que nous proposons, nous gardons l'équivalent logarithmique, c'est à dire (index l supprimé pour des raisons de clareté) :

$$\mathbb{P}_L \asymp e^{(-NI_1(c,0)-\theta_1^* B_o(N))} \tag{6.14}$$

Le point de travail s^* est solution de l'équation :

$$b+ct-\frac{mpte^{spht}}{1-\frac{m}{h}+\frac{m}{h}e^{spht}} = 0$$

donc

$$s^* = \frac{1}{pht}\log\left(\frac{(m-h)(b+ct)}{m(b+(c-ph)t)}\right)$$

Les cas qui nous intéressent sont (afin de garantir la stabilité du système et qu'il y ait des pertes)

$$\frac{b+ct}{b+(c-hp)t)} > \frac{h}{h-m}$$
$$c > mp - \frac{b}{t}$$
$$c < hp - \frac{b}{t}$$

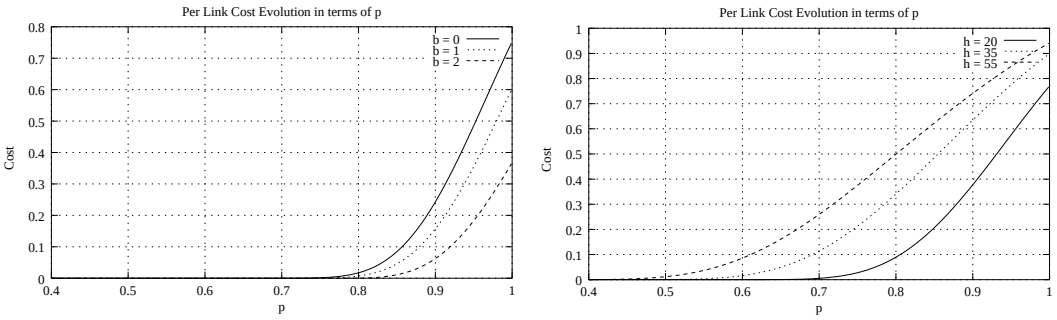
La fonction de taux résultante est donnée par l'expression :

$$I_t(b+ct) = \frac{(b+ct)}{pht}\log\left(\frac{(m-h)(b+ct)}{m(b+ct-pht)}\right) - \log\left(\frac{pt(m-h)}{b+ct-pht}\right)$$

Sous l'hypothèse de petit buffer, nous calculons $I_1(c)$ et nous dérivons la fonction de coût par lien :

$$\mathrm{K} = p \times e^{\left\{-N\left(\frac{c}{hp}\log\left(\frac{c(m-h)}{m(c-ph)}\right)-\log\left(\frac{p(m-h)}{c-ph}\right)\right)-s^*|_{\substack{b=0\\t=1}}b\sqrt{N}\right\}} \tag{6.15}$$

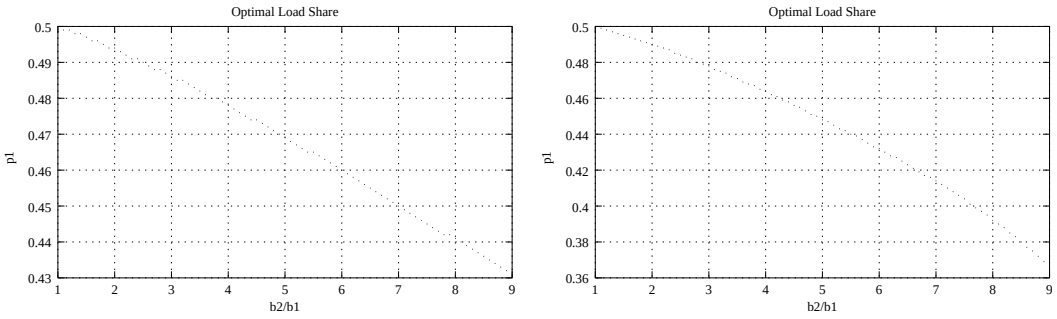
Une fois que nous avons obtenu l'expression de la fonction de coût par lien, il est possible de résoudre numériquement des instances du problème. Comme exemple, la fonction de coût par lien, en fonction



(a) Taille de buffer variable.

(b) Débit crête variable variable.

Fig. 6.13: Illustration de la fonction de coût par lien (partage x équivalent logarithmique) en fonction de la proportion de trafic p acheminé sur le lien. Le trafic d'entrée est l'agrégation de $N = 30$ sources caractérisées par leur b.p.e. de Hoeffding ($m = 9$) avec une capacité par source $c = 10$. Nous pouvons remarquer que les fonctions de coût par lien ne sont pas strictement convexes.

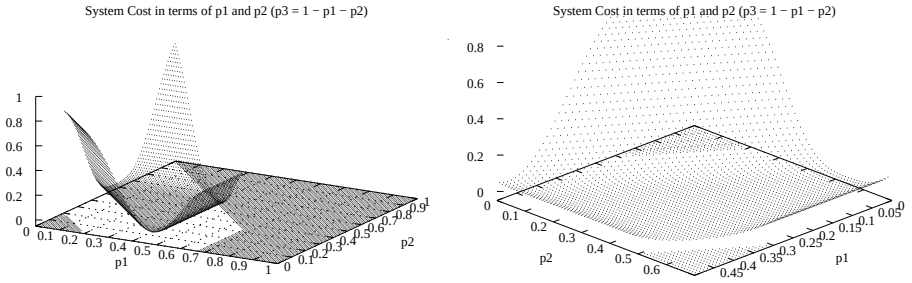


(a) $h = 45$

(b) $h = 30$

Fig. 6.14: Evolution du partage de charge optimal dans un système à 2 liens, symétrique où le trafic d'entrée est l'agrégation de $N = 30$ sources caractérisées par leur b.p.e. de Hoeffding ($m = 10, h = 45$) et (1) $C_1 = N10, B_1 = \sqrt{N}1$ (2) $C_2 = N10, B_2 = \sqrt{N}b_2$. Les figures montrent l'évolution du partage (avec $B_1 = B_2$ nous attendons un partage de $0.5, 0.5$). Notons l'effet des différents rapports de taille de buffer par source.

de la proportion de trafic p acheminé sur le lien est illustré dans la figure [**?**] où le trafic d'entrée est l'agrégation de $N = 30$ sources caractérisées par leur b.p.e. de Hoeffding ($m = 9$) avec une capacité par source $c = 10$ (unités de travail produit/servi par unité de temps). Nous pouvons remarquer que les fonctions de coût par lien ne sont pas strictement convexes. Ensuite, le partage de charge optimal est calculé dans un système à 2 liens, symétrique où le trafic d'entrée est l'agrégation de $N = 30$ sources caractérisées par leur b.p.e. de Hoeffding ($m = 10, h = 45$) et (1) $C_1 = N10, B_1 = \sqrt{N}1$ (2) $C_2 = N10, B_2 = \sqrt{N}b_2$. Les figures montrent l'évolution du partage (avec $B_1 = B_2$ nous attendons



(a) Général (b) Détail

Fig. 6.15: Fonction objective pour un système à 3 liens, avec trafic d'entrée agrégation de $N = 40$ sources caractérisées par leur débit moyen ($m = 15$) et leur débit crête ($h = 45$), utilisant la borne supérieure de la b.p.e. Les liens sont modélisés comme une file d'attente avec (1) $C_1 = N10, B_1 = \sqrt{N}4$, (2) $C_2 = N15, B_2 = \sqrt{N}2$ et (3) $C_3 = N12, B_3 = \sqrt{N}1$. Le partage optimal correspond au vecteur $(0.31, 0.39, 0.30)$. Remarquons dans la figure (a) les zones rayées correspondant à des combinaisons ne respectant pas les contraintes de stabilité. Dans la figure (b), remarquons le caractère assez plat autour d'un voisinage de l'optimum.

un partage de $0.5, 0.5$). Notons l'effet des différents rapports de taille de buffer par source (cf. 6.14).

Avec ces exemples, nous avons mis en valeur l'intérêt d'une approche pour le calcul du partage de charge qui prend en compte les propriétés stochastiques du trafic et les différentes contraintes de délai et de taux de pertes applicables a des routes différentes (par exemple, les différents seuils dans le dernier exemple).

L'analyse détaillé de ces cas nous a permis d'identifier une famille de problèmes d'optimisation dont la solution est invariante et peut être généralisée, que nous présentons dans la suite.

6.7. Généralisation du partage de charge

Notons que certains résultats présentés peuvent être généralisés, d'après le théorème suivant :

Théorème 6.7.1 (Partage Proportionnel). *Soit $f(x)$ une fonction continue, positive, strictement convexe et croissante. Le problème d'optimisation non linéaire de la forme :*

$$\min_{p \in \mathbb{R}^L} F(p) = \min_{p \in \mathbb{R}^L} \sum_{l=1}^{L} p_l f\left(\frac{p_l}{K_l}\right) \tag{6.16}$$

avec les contraintes :

$$0 < p_l \leq 1 \qquad \forall l \in 1 \dots L$$
$$\sum_{l=1}^{L} p_l = 1$$

admet un minimum global contraint donné par l'expression :

$$p_i^* = \frac{K_i}{\sum_{l=1}^{L} K_l} \tag{6.17}$$

Démonstration. La démonstration est similaire à celle utilisée pour trouver les conditions d'optimalité. La matrice Hessiene est définie positive. Formalisons les contraintes :

$$\begin{aligned} g_k(p) &= p_k \geq 0 \qquad k = 1 \dots L \\ c_m(p) &= 1 - p_k \geq 0 \qquad m = 1 \dots L \\ h(p) &= \sum_{l=1}^{L} p_l - 1 = 0 \end{aligned}$$

Remarquons que les contraintes d'égalité et les contraintes d'inégalité sont linéaires. Le Lagrangien est donné par :

$$L(p,u,w,v) = \sum_{l=1}^{L} p_l f\left(\frac{p_l}{K_l}\right) - \sum_{l=1}^{L} g_l(p)u_l - \sum_{l=1}^{L} c_l(p)w_l - v\left(\sum_{l=1}^{L} p_l - 1\right)$$

Le calcul des dérivées partielles et les conditions de Karush-Kuhn-Tucker, s'écrivent

$$\begin{aligned} \frac{\partial L(p,u,w,v)}{\partial p_i} &= f\left(\frac{p_i}{K_i}\right) + Df\left(\frac{p_i}{K_i}\right)\frac{p_i}{K_i} - u_i + w_i - v = 0 \\ \frac{\partial g_l(p)}{\partial p_i} &= 1_{\{l=i\}} \\ \frac{\partial c_l(p)}{\partial p_i} &= -1_{\{l=i\}} \\ \frac{\partial h(p)}{\partial p_i} &= 1 \end{aligned}$$

Supposons les contraintes d'inégalité inactives, $w_i = 0$ et $u_i = 0$, alors :

$$f\left(\frac{p_i}{K_i}\right) + Df\left(\frac{p_i}{K_i}\right)\frac{p_i}{K_i} = v \qquad \forall i = 1 \dots L$$

$$\forall i,j \; f\left(\frac{p_i}{K_i}\right) + Df\left(\frac{p_i}{K_i}\right)\frac{p_i}{K_i} = f\left(\frac{p_j}{K_j}\right) + Df\left(\frac{p_j}{K_j}\right)\frac{p_j}{K_j} \tag{6.18}$$

Etant donné que f est strictement croissante et convexe, alors $f(x) + xDf(x)$ est strictement croissante et l'équation (6.18) est vérifiée si et seulement si $\frac{p_i}{K_i} = \frac{p_j}{K_j}$. La solution optimale respectant les contraintes est :

$$p_i^* = \frac{K_i}{\sum_{l=1}^{L} K_l} \tag{6.19}$$

Les conditions de Karush-Kuhn-Tucker de premier ordre sont vérifiées :

$\forall i = 1 \dots L$

$$\begin{aligned} g_i(p^*) &\geq 0 \\ c_i(p^*) &\geq 0 \\ h(p^*) &= 0 \\ w_i c_i(p^*) &= 0 \quad w_i \geq 0 \\ u_i g_i(p^*) &= 0 \quad u_i \geq 0 \end{aligned}$$

□

Egalisation de la charge

On remarquera que les fonctions de coût proposées (voir par exemple les cas avec des sources fBm ou Poisson) peuvent être écrites sous la forme citée *à condition que tous les seuils / tailles de buffer puissent s'écrire comme* $B_i = DC_i$. Si la convexité par rapport à p est démontrée on trouve des conditions d'optimalité globale, et le partage de charge optimal est proportionnel à la capacité nominale (résiduelle si le trafic externe est constant). Ceci implique que les partages de charge sont calculés de telle sorte que les charges de chaque file d'attente sont égales.

6.8. Partage «orienté connexion»

L'approche fluide proposée peut s'avérer difficilement applicable. En conséquence, dans cette section nous allons évaluer rapidement une deuxième approche, complémentaire à celle proposée dans ce chapitre, et qui appartiendrait plutôt à la classe «partage orienté connexion» (cf.section 4.10.3) : Nous supposons que l'agrégat de trafic est composé d'un nombre N_T de sources i.i.d, et que le partage de charge est réalisé en choisissant un vecteur $\mathbf{N} = (N_1, \dots, N_L)$, vérifiant $\sum_{l=1}^{L} N_l = N_T$ et $N_l \in I\!N, \forall l = 1..L$. Chaque composante N_l du vecteur correspond au nombre de sources à *router* en tant que flots sur la file d'attente l, afin de respecter un certain critère. Nous pouvons citer entre autres :

1. Minimisation du volume moyen de pertes. Ceci est illustré dans la suite.
2. Minimisation du coût marginal. Autrement dit, la répartition des sources parmi les files d'attente est telle que le fait d'ajouter une nouvelle source à l'agrégat amène une augmentation minimale d'un critère de performance tel que le taux de pertes ou la probabilité de débordement.

Un exemple

Comme exemple, considérons le problème d'optimisation suivant :

$$\min_{\mathbf{N}} \sum_{l=1}^{L} N_l \lambda L R_l \tag{6.20}$$

où

$$\begin{aligned} \lambda \triangleq \lim_{s\to 0} \alpha(s,t) \quad & \text{est le débit moyen d'une source de l'agrégat et} \\ LR_l \quad & \text{est le taux de pertes de la file } l \end{aligned} \tag{6.21}$$

avec les contraintes :

$$\begin{aligned} \sum_{l=1}^{L} N_l = N_T \quad & \text{Contrainte de partage} \\ N_l \lim_{s\to 0} < C_l \quad & \text{Contrainte de stabilité} \\ 0 \leq N_l \leq N_T \forall l = 1..L & \\ N_l \in I\!N \forall l = 1..L & \end{aligned} \tag{6.22}$$

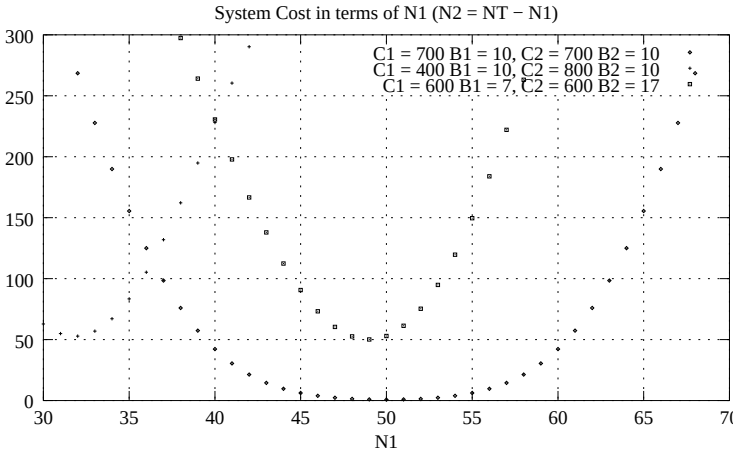


Fig. 6.16: Répartition optimale du trafic avec L=2 en fonction du nombre de sources routées sur le lien 1, avec $N_T = 100$ sources de «hoeffding», m=9, et h=20

La figure 6.16 illustre la fonction objective en fonction du nombre de sources N_1 «routées» sur le lien 1 (donc $N_2 = N_T - N_1$). Nous utilisons la borne de Hoeffding, avec $m = 9, h = 20$, pour un agrégat d'entrée composé de $N_T = 100$ sources i.i.d, avec une topologie à $L = 2$ liens, pour différentes combinaisons de capacités et de tailles de buffer C_l, B_l. Le taux de pertes est calculé avec l'asymptotique du grand nombre d'usagers, avec $c_l = C_l/N_l$ et $b_l = B_l/\sqrt{N_l}$. Comme prévu, des liens avec une plus grande capacité acceptent un plus grand nombre de sources, et à capacité égale, c'est la taille du buffer qui influence la répartition de trafic.

Commentaires

1. Nous avons illustré cette deuxième approche avec des modèles de trafic utilisant la borne de Hoeffding. Plusieurs classes de trafic (en respectant la discipline FIFO) peuvent être aussi supportées en choisissant *une matrice,* dont les composantes correspondent au nombre de

sources par file et par classe de trafic.

2. Le caractère *discret* des valeurs N_l rend (encore plus) difficile la résolution du problème. Néanmoins, d'après la figure 6.16, il paraît possible d'appliquer des techniques de relaxation aux réels.
3. Rappelons que la probabilité de débordement et le taux de pertes sont approximées en utilisant l'asymptotique, qui nécessite que le nombre de sources multiplexées soit «assez grand» pour être raisonnablement précise. En pratique, ceci veut dire qu'il est nécessaire d'imposer de nouvelles contraintes sur les valeurs de type $N_l \geq N_{\min}$, où $N_{\min}$ est une valeur entière, arbitrairement choisie, considérée comme la plus petite valeur à partir de laquelle l'approximation est acceptable. Nous estimons que des valeurs de quelques dizaines de sources sont suffisantes.
4. L'identification formelle du rapport entre les vecteur $\mathbf{p}$, variable de décision présentée dans ce chapitre, et $\mathbf{N}$ reste encore une question ouverte. Dans certains cas nous pouvons affirmer que $N_l^* \approx p_l^* N_T$.

6.9. Remarques et Conclusions

Remarques Finales

– Notons que même dans les cas relativement simples que nous avons présentés, il paraît difficile d'obtenir des expressions explicites pour les quantités qui nous intéressent. Citons par exemple le point de travail t^* du deuxième exemple : sa valeur est donnée sous la forme d'une équation implicite nécessitant le calcul des dérivées implicites.
– Il est relativement simple d'étendre la fonction objective en utilisant les expressions asymptotiquement exactes obtenues dans [58], au lieu des équivalents logarithmiques utilisés. Nous avons choisi cette deuxième approche pour sa relative simplicité par rapport à la première, dans laquelle il s'avère très difficile d'obtenir des conditions d'optimalité. A l'exception du terme de variance, le calcul des coefficients apparaissant devant les exponentielles sont fonction de paramètres comme le point de travail spatial, le nombre de sources multiplexées ou le rapport entre le débit moyen d'une source et la capacité par source.

Synthèse des contributions

Dans ce chapitre, nous avons proposé un modèle pour le partage de charge sur une topologie multi-lien. L'étude était conçue pour l'architecture MPLS, mais la plupart des résultats peuvent être appliqués à d'autres contextes. De façon générale, le partage de charge optimal dépend des propriétés statistiques de l'agrégat de trafic. L'approche proposée est légèrement pénalisée par la complexité des calculs intervenant dans les problèmes d'optimisation, mais certains auteurs ont proposé des heuristiques pour résoudre la double optimisation qui apparaît dans le calcul des fonctions de taux. Avec les calculateurs actuels, pour des systèmes de taille petite ou moyenne, une simple approche de recherche exhaustive peut être acceptable.

– Dans le modèle proposé, nous n'imposons pas de limites ou d'hypothèses non réalistes. La caractérisation du trafic d'entrée est réalisée au moyen des bandes passantes effectives, avec des hypothèses assez génériques d'ergodicité et de stationnarité.

– Nous utilisons une approche fluide pour la caractérisation du trafic offert à chaque lien. Ainsi, les b.p.e. offertes sont une version amincie *(ang. thinned)* correspondant à un changement d'échelle en p et en s (le paramètre spatial). Ceci présente l'avantage qu'il est facile de déterminer des expressions pour les bandes passantes effectives offertes à partir des b.p.e. d'entrée.

Résultats

Nous avons étudié un critère particulier consistant à minimiser le *Taux de Pertes/Overflow.* Il est important de noter que la probabilité *de dépasser un seuil* dans une file à buffer infini et *le taux de pertes* ont des équivalents logarithmiques égaux (même fonction de taux), ce qui nous permet d'unifier les deux critères. Ainsi, dans un système de tailles de buffers finies, nous pouvons voir la fonction de coût comme le taux de pertes global du système. Bien entendu, il est toujours possible de proposer d'autres critères de performance, et l'on pourrait appliquer des techniques similaires. Pour ce critère, nous avons obtenu des conditions d'optimalité et nous avons illustré l'approche avec des exemples significatifs.

– Les fonctions de coût (et la performance du système dans des état sous optimaux) dépendent des propriétés de trafic. Néanmoins, dans certaines configurations particulières dans lesquelles les seuils sont proportionnels aux capacités, le partage de charge optimal est insensible aux propriétés du trafic et correspond à des règles simples et pragmatiques comme *proportionnalité par rapport à la capacité résiduelle,* qui amènent à l'égalisation de la charge.
– Nos études permettent l'obtention du partage de charge optimal et l'évaluation de la dégradation des performances quand le système travaille dans des états non optimaux (par exemple, suite à une panne ou dans le cas ou l'on travaille avec des informations de routage obsolètes). Il suffit de comparer les valeurs de la fonction de coût dans l'optimum et dans l'état non optimal.

Dans les chapitres suivants, nous proposons des modèles qui essaient de lever certaines limitations du modèle présenté ici, notamment en introduisant la variabilité temporelle de la capacité d'un LSP et en proposant des approches qui généralisent les résultats obtenus à un réseau dans sa globalité.

7. Partage de charge à Capacité variable

Extensions de bout en bout du partage de charge

7.1. Introduction

Dans le chapitre précédent, nous avons obtenu des conditions d'optimalité pour le partage de charge lorsque le système peut être modélisé par une topologie multi-lien. Les résultats obtenus peuvent être étendus de manière directe à un contexte de bout en bout [1] si l'on fait l'hypothèse que la capacité associée à chaque tunnel est issue d'une réservation stricte de ressources. Rappelons que l'optimalité du partage est *locale* au routeur d'entrée.

7.1.1. Motivation

L'objectif de ce chapitre est d'incorporer au modèle une notion de bout en bout sous l'hypothèse d'un modèle de service «uniclasse» (sans distinction de classes de trafic), dans lequel les LSPs sont établis sans réservation stricte de ressources, toujours dans la perspective d'un partage optimal local. Une caractéristique notable de ce système est *la variabilité temporelle* de la capacité associée à chaque LSP, due principalement aux multiplexages statistiques ayant lieu dans les liens traversés par le LSP.

Dans ce cadre, plusieurs approches de modélisation sont possibles et seront présentées dans la suite. L'idée de base est d'intégrer les phénomènes aléatoires de la capacité résiduelle vue du routeur d'entrée.

Solution Statique

Le partage de charge est calculé selon un critère statique, en estimant un coût pour chaque LSP, en fonction des attributs statiques des liens qui le composent, comme par exemple, les poids administratifs associés. Cette approche ne prend pas en compte les propriétés statistiques du trafic.

Solution Adaptative

Cette approche consiste à développer l'approche présentée au chapitre précédent de façon adaptative. La bande passante associée a chaque LSPs est calculée suite à la réception des messages d'états de lien *(LSA Link State Advertisement)* . Typiquement, on prend comme capacité nominale du LSP la plus petite capacité résiduelle des liens formant le LSP. Bien sûr, cette approche nécessite l'hypothèse de l'existence d'un protocole de routage sensible à la qualité de service, capable d'assurer la distribution de l'attribut *capacité résiduelle* de chaque lien. L'algorithme devient adaptatif en recalculant à chaque mise à jour le nouveau partage de charge. Les principaux inconvénients de cette approche sont la

[1] Au sens d'un domaine MPLS, entre un routeur d'entrée et un routeur de sortie

charge importante que ceci impose au CPU et les phénomènes possibles d'oscillation et d'instabilité, associés aux délais de distribution d'information du protocole de routage.

Solution Stationnaire

L'approche présentée dans ce chapitre consiste à caractériser l'évolution au cours du temps de *la capacité associée à chaque LSP* en considérant un processus aléatoire, avec un nombre plus ou moins limité de paramètres. En gros, la modélisation à l'aide de processus stochastiques tente de prendre en compte la nature des variations temporelles, et de les incorporer aux problèmes d'optimisation proposés, qui sont étendus en considérant ce que l'on définit comme *Capacité Effective,* terme qui sera défini et formalisé dans les sections suivantes. Nous évoquons à la fin de ce chapitre une approche adaptative visant notamment à pallier le caractère non stationnaire des systèmes réels. En ce sens, le modèle présenté dans le chapitre chapitre 6, *Partage de charge sur une topologie multi-lien*, peut être vu comme un cas particulier du modèle présenté dans la suite, pour lequel la capacité apparaissant dans les fonctions de coût correspond à une capacité fixe.

La statistique des processus modélisant la capacité de chaque LSP vue du routeur d'entrée est supposée connue, et issue par exemple de techniques de métrologie.

7.2. Modèle du Système

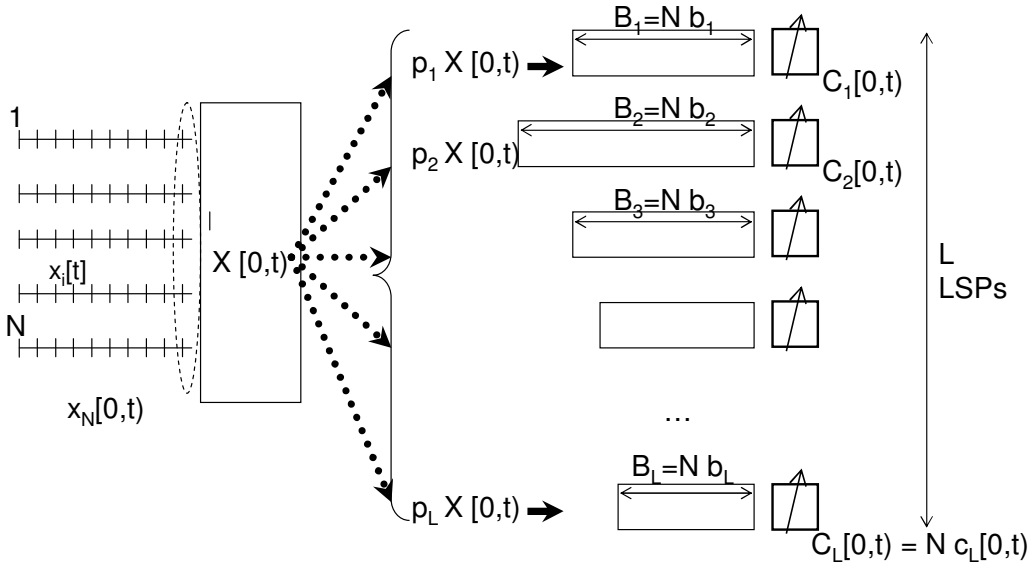


Fig. 7.1: Modèle du Système

Le modèle est très similaire à celui que nous avons présenté dans l'analyse de la topologie multi-lien et est illustré sur la figure 7.1. Considérons un système à temps discret, indexé par $t \in \mathbb{N}$. La

description du système sera réalisée en trois étapes : d'abord, la caractérisation du trafic d'entrée, ensuite la caractérisation du trafic offert à chaque LSP et finalement, la caractérisation de chaque LSP appartenant au groupe de LSPs.

Caractérisation et Notation du Trafic d'Entrée

Le trafic d'entrée est caractérisé par un processus stochastique cumulatif $X_t^{(N)}$ et est constitué d'un agrégat de N sources i.i.d. La notation utilisée est la suivante :

$$\begin{aligned} x_i[t] &\rightarrow \text{travail produit par une source (source i) à l'instant t} \\ x_i[0,t) = \sum_{m=0}^{t-1} x_i[m] &\rightarrow \text{travail total produit par une source pendant l'intervalle [0,t)} \\ X[t] = \sum_{i=1}^{N} x_i[t] &\rightarrow \text{travail produit par l'agrégat à l'instant t} \\ X[0,t) = X_t^{(N)} = \sum_{i=1}^{N} \sum_{m=0}^{t-1} x_i[m] &\rightarrow \text{travail total produit par l'agrégat pendant l'intervalle [0,t)} \end{aligned} \tag{7.1}$$

Caractérisation du trafic offert par LSP

Nous reprenons la caractérisation utilisée aux chapitres précédents : le trafic offert à chaque LSP l sera déterminé par le trafic agrégé d'entrée $X_t^{(N)}$ et par un paramètre $p_l \in [0,1]$, nommé *partage associé au LSPl*. Ainsi, le trafic total offert au LSP l pendant l'intervalle $[0,t)$ sera noté $p_l X[0,t)$ ou simplement $X_{t,l}^{(N)} \triangleq p_l X_t^{(N)}$.$p_l$ correspond à la proportion fluide de travail par source acheminé sur le LSP l. On a alors :

$$\begin{aligned} p_l x_i[t] &\rightarrow \text{travail produit par la source i à l'instant t et acheminé sur le lien } l \\ p_l x_i[0,t) = \sum_{m=0}^{t-1} p_l x_i[m] &\rightarrow \text{travail total produit par une source pendant [0,t) et acheminé sur le lien } l \\ X_{t,l}^{(N)} = p_l X_t^{(N)} &\rightarrow \text{travail total offert à la file } l \text{ pendant l'intervalle [0,t)} \end{aligned} \tag{7.2}$$

Caractérisation des LSPs

Chaque LSPs est modélisé par une file d'attente et indexé par $l \in \{1..L\}$. La taille du buffer associé à la file d'attente est noté $B_l = Nb_l$, c'est-à-dire que la taille du buffer croît linéairement avec le nombre de sources constituant l'agrégat de trafic. Les différents processus de service sont supposés

indépendants de l'agrégat de trafic. La notation utilisée est la suivante :

$$\begin{aligned} C_l^{(N)}[t] &= Nc_l[t] \rightarrow \text{Travail total servi par le serveur de la file } l \text{ à l'instant t} \\ C_{t,l}^{(N)} \triangleq C_l[0,t) &= \sum_{m=0}^{t-1} Nc_l[m] \rightarrow \text{Travail total servi par le serveur de la file } l \text{ pendant [0,t)} \\ c_l[0,t) &= \frac{C_l[0,t)}{N} \rightarrow \text{idem, normalisé par le nombre de sources de l'agrégat} \end{aligned} \tag{7.3}$$

On introduit la notion de *Trafic Résiduel par source et par agrégat.* Pour chaque file l :

$$\begin{aligned} r_i^l[0,t) &= \sum_{m=0}^{t-1} (p_l x_i[m] - c_l[m]) \\ R_{t,l}^{(N)} &= R_l^{(N)}[0,t) = \sum_{i=1}^{N} r_i^l[0,t) \\ R_{t,l}^{(N)} &= p_l X_t^{(N)} - C_{t,l}^{(N)} = X_{t,l}^{(N)} - C_{t,l}^{(N)} \end{aligned}$$

7.2.1. Equation de Lindley

L'équation de Lindley décrivant l'évolution temporelle du travail cumulé dans une file d'attente (contexte discret) admet une extension directe au cas d'un système à capacité variable. Les expressions suivantes sont essentielles pour obtenir les asymptotiques que nous utilisons.

Equation de Lindley pour des systèmes à buffer infini

$$W_l{}^{(N)}[t] = \max\left(0, W_l{}^{(N)}[t-1] + p_l X^{(N)}[t] - C_l{}^{(N)}[t]\right) \tag{7.4}$$

Equation de Lindley pour des systèmes à buffer fini

$$Y_l{}^{(N)}[t] = \min\left(\max\left(0, Y_l^{(N)}[t-1] + p_l X^{(N)}[t] - C_l{}^{(N)}[t]\right), Nb_l\right) \tag{7.5}$$

Dans la suite, nous analysons une unique file d'attente isolée. Nous commençons par caractériser la queue de distribution du processus de trafic offert résiduel, en appliquant directement le théorème de Cramer. En utilisant les mêmes arguments que Likhanov et Mazumdar [58], il est possible d'obtenir des expressions pour les asymptotiques de la queue de distribution du travail cumulé dans le cas à buffer infini ou du taux de pertes. A partir des équivalents logarithmiques obtenus, nous formulons le problème du partage de charge à capacité variable comme un problème d'optimisation non linéaire sous contraintes.

7.2.2. Queue de Distribution du processus de Trafic Résiduel

Définition 7.2.1 (Ordre Stochastique).

$$X[0,t) \leq_{st} C[0,t) \Longleftrightarrow \mathbb{P}[X[0,t) \leq \mu] \leq \mathbb{P}[C[0,t) \leq \mu] \quad \forall \mu \tag{7.6}$$

Proposition 7.2.1 (Comparaison Stochastique). *Soient $X[0,t)$ et $C[0,t)$ deux processus stochastiques à valeurs dans $\mathbb{R}^+$. Soit $f : \mathbb{R}^+ \to \mathbb{R}^+$, une fonction mesurable et croissante. Si $X[0,t) \leq_{st} C[0,t)$, alors :*

$$\log \mathbb{E}[e^{\theta C[0,t)}] \geq \log \mathbb{E}[e^{\theta X[0,t)}] \qquad \forall t, \theta \geq 0 \tag{7.7}$$

Démonstration. Notons d'abord que

$$\mathbb{E}[f(X[0,t))] = \int_{\mathbb{R}^+} \mathbb{P}(f(X[0,t) \geq \mu) d\mu$$

$$\begin{aligned} X[0,t) \leq_{st} C[0,t) &\Longrightarrow \mathbb{E}[f(X[0,t))] \leq \mathbb{E}[f(C[0,t))] \\ &\Longrightarrow \log \mathbb{E}[f(X[0,t))] \leq \log \mathbb{E}[f(C[0,t))] \\ &\Longrightarrow \log \mathbb{E}[e^{\theta X[0,t)}] \leq \log \mathbb{E}[e^{\theta C[0,t)}] \forall t, \theta \geq 0 \end{aligned}$$

□

La caractérisation de la queue de distribution du processus de trafic résiduel fait intervenir les transformées convexes des log Laplaces des processus définis précédemment. On les note :

$\Lambda_t^X(\theta_t) \triangleq \log \mathbb{E}[e^{\theta_t x_1[0,t)}]$ T. LogLaplace du processus de travail produit par une source

$\Lambda_{t,l}^X(\theta_t) \triangleq \log \mathbb{E}[e^{\theta_t p_l x_1[0,t)}]$ T. LogLaplace du processus de travail produit par une source et offert au LSP l

$\Lambda_{t,l}^K(\theta_t) \triangleq \log \mathbb{E}[e^{\theta_t C_l[0,t)}]$ T. LogLaplace du processus de service du LSP l

$\Lambda_{t,l}^C(\theta_t) \triangleq \log \mathbb{E}[e^{\theta_t c[0,t)}] = \Lambda_{t,l}^K(\frac{\theta_t}{N})$ T. LogLaplace du processus de service l par source

$\Lambda_{t,l}^R(\theta_t) \triangleq \Lambda_{t,l}^X(\theta_t) + \Lambda_{t,l}^C(-\theta_t)$ T. LogLaplace du processus de trafic résiduel

$I_{t,l}^R(b_l) \triangleq \sup_\theta \left\{ b_l \theta_t - \Lambda_{t,l}^R(\theta_t) \right\}$ Fonction de Taux associée au procesus $r_l[0,t)$

Hypothèse de stabilité

Dans le contexte du partage de charge, cette hypothèse s'applique à toutes les files d'attente composant le groupe de LSPs et concerne les moyennes des processus de trafic offerts et la capacité moyenne du LSP. C'est une condition suffisante de stabilité impliquant l'existance d'une distribution stationnaire du travail cumulé dans les files d'attente.

$$\frac{\partial \Lambda_{t,l}^{R}(\theta_t)}{\partial \theta_t} |_{\theta_t=0} < 0 \quad \Leftrightarrow \quad \mathbb{E}[p_l x[0,t)] < \mathbb{E}[c_l[0,t)] \tag{7.8}$$

Nous verrons que cette hypothèse impose une famille de contraintes au problème d'optimisation que nous proposerons par la suite.

L'événement «débordement» correspond à la caractérisation de la queue de distribution du trafic résiduel de la file d'attente.

$$\begin{aligned}
\mathbb{P}\left(X_{t,l}^{(N)} - C_{t,l}^{(N)} > Nb_l\right) &= \mathbb{P}\left(\sum_{i=1}^{N}\sum_{n=0}^{t-1} p_l x_i[n] - N\sum_{n=0}^{t-1} c_l[n] > Nb_l\right) \\
&= \mathbb{P}\left(\frac{\sum_{i=1}^{N}\sum_{n=0}^{t-1} p_l x_i[n]}{N} - N\frac{\sum_{n=0}^{t-1} c_l[n]}{N} > b_l\right) \\
&= \mathbb{P}\left(\frac{\sum_{i=1}^{N} r_i^l[0,t)}{N} > b_l\right)
\end{aligned} \tag{7.9}$$

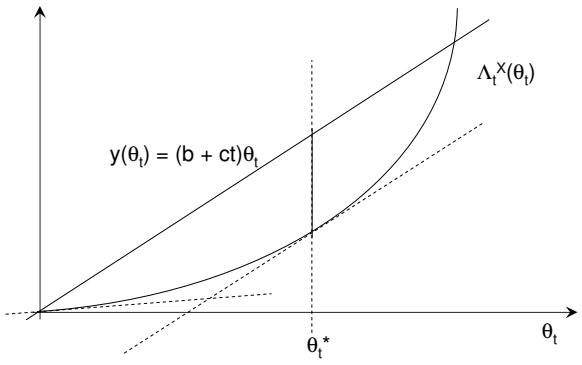
La queue de distribution du trafic résiduel, sous les hypothèses mentionnées, est obtenue en appliquant le théorème de Bahadur Rao

$$\mathbb{P}\left(R_{t,l}^{(N)} > Nb_l\right) = \frac{1}{\theta_t^* \sqrt{2\pi\sigma_t^2 N}} e^{-NI_{t,l}^{R}(b_l)} \left(1 + O\left(\frac{1}{N}\right)\right)$$

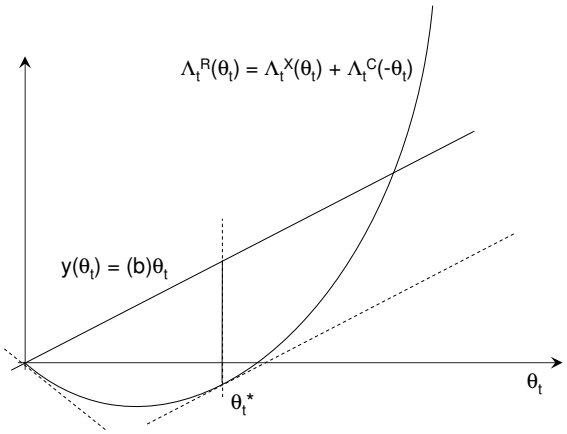
où θ_t^* est appelé *le point spatial de travail* et correspond à la valeur qui réalise le maximum de la fonction de taux, et où σ_t^2 est obtenu comme indiqué dans la section 5.8.3.1. Nous verrons par la suite que les critères de performance que nous utilisons font intervenir cette fonction de taux. Si l'on néglige le terme apparaissant devant l'exponentiel, on parle d'*équivalents logarithmiques.*

Remarque :

Dans un contexte de capacité fixe, il est possible de démontrer l'existence et l'unicité de la valeur θ^* réalisant le supremum apparaissant dans la fonction de taux (terme $I_t^R(b)$). Ceci est illustré par la figure 7.2 car la différence entre la droite $y(\theta_t) = (b + ct)\theta_t$ et la fonction $\Lambda_t^X(\theta_t)$ est concave (et unimodale, car $\Lambda_t^X(\theta_t)$ est convexe, figure (a)). A droite, les mêmes arguments s'appliquent aux extensions à capacité variable (figure (b)). Remarquons que dans le cas à capacité fixe, la pente de la courbe Λ en $\theta = 0$ est déterminée par l'espérance du processus de trafic. Dans le cas à capacité variable cette pente est négative (hypothèse 7.8) et est égale à $\mathbb{E}[x_1[0,t)] - \mathbb{E}[c[0,t)]$.



(a) Capacité fixe.



(b) Capacité variable.

Fig. 7.2: Existance et Unicité de θ_t^*

7.2.3. Asymptotique du grand nombre d'usagers

Systèmes à buffer infini

La queue de distribution du travail cumulé (stationnaire) dans une file d'attente est donnée par l'expression :

$$\mathbb{P}(W_l^{(N)} > Nb_l) \approx e^{-NI_{t^*}^R(b_l)}$$

Systèmes à buffer fini

L'équivalent logarithmique du taux de pertes est donné par :

$$\mathbb{P}_l = \frac{\mathbb{E}\left[\max\left(0, Y^{(N)}[t-1] + \sum_{i=1}^{N} p_l x_i[t] - Nb_l - Nc_l[t]\right)\right]}{\mathbb{E}[X_{t,l}^{(N)}]}$$
$$\approx e^{-NI^R_{t^*,l}(b_l)}$$

où t^* correspond à la valeur qui minimise $I_t^R(b)$. Dans notre cas, nous allons utiliser l'équivalent logarithmique uniquement, nous permettant d'unifier les systèmes à buffer infini (probabilité de dépasser un seuil) et les systèmes à buffer fini (probabilité de saturation, avec le même équivalent logarithmique que le taux de pertes).

7.2.4. La notion de «capacité effective»

Plusieurs auteurs utilisent la notation introduite par Kelly [53] concernant les bandes passantes effectives. De la même manière, nous introduisons le terme de «capacité effective» définie par :

$$\begin{aligned}\kappa(s,t) &\triangleq \frac{-\Lambda_t^c(-s)}{st} \\ &= -(st)^{-1}\log\mathbb{E}[e^{-sc[0,t)}]\end{aligned} \tag{7.10}$$

Propriétés de la capacité effective

- La capacité effective est décroissante avec s.
- Quand $s \to 0$, la capacité effective associée au processus de service tend vers sa capacité moyenne :

$$\lim_{s\to 0}\kappa(s,t) = \frac{\mathbb{E}[c[0,t)]}{t} \tag{7.11}$$

- Quand $s \to \infty$, la capacité effective associée au processus de service tend vers sa capacité minimale :

$$\lim_{s\to\infty}\kappa(s,t) = \frac{\operatorname{ess\,inf}(c[0,t))}{t} \tag{7.12}$$

Extension de la formule «infsup»

Le terme «infsup», normalement défini comme

$$-\inf_t \sup_s \{(b+ct)s - st\alpha(s,t)\} \tag{7.13}$$

devient, en appliquant la notion de capacité effective[2]

$$-\inf_t \sup_s \{bs + st\kappa(s,t) - st\alpha(s,t)\} \tag{7.14}$$

7.3. Approximations

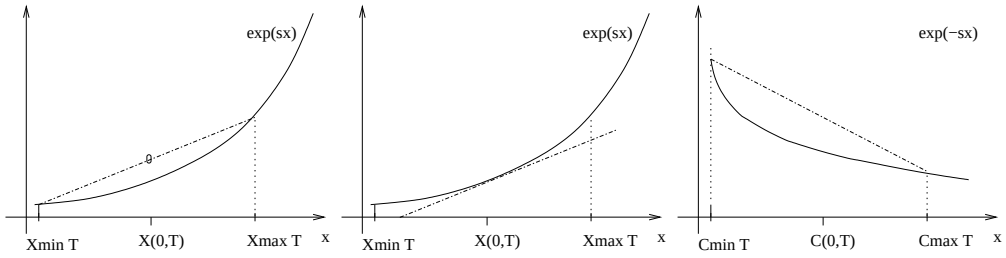
Il est parfois difficile de caractériser les transformées de Laplace des processus intervenant dans les différents problèmes d'optimisation, leurs formes analytiques présentant une certaine complexité,

[2] un cas particulier est alors $\kappa(s,t) = c$.

ou étant tout simplement inconnues. Comme évoqué dans les chapitres précédents, en utilisant les propriétés des fonctions convexes, il est possible d'obtenir des bornes sur les bandes passantes effectives ainsi que sur les capacités effectives, dans notre cas les fonctions $x \rightarrow exp(sx)$ et $x \rightarrow exp(-sx)$, $s \geq 0$:

Proposition 7.3.1 (propriété des fonctions convexes). *Soit $f(x)$ une fonction mesurable, convexe dans l'intervale $[a, b]$. Pour tout $x \in [a, b]$:*

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) \tag{7.15}$$



(a) e^{sx} (b) e^{sx} (c) e^{-sx}

Fig. 7.3: Capacités Effectives : Bornes Linéaires

Borne Supérieure de la bande passante effective

(cf. figure 7.3 (a)). Si $x[t] \in [x_{\min}, x_{\max}]$, et $x_{\text{avg}} \triangleq \frac{\mathbb{E}[X[0,t)}{t}$:

$$\alpha(s,t) \leq \frac{1}{st} \log \left(\frac{x_{\max} - x_{\text{avg}}}{x_{\max} - x_{\min}} e^{s x_{\min} t} + \frac{x_{\text{avg}} - x_{\min}}{x_{\max} - x_{\min}} e^{s x_{\max} t} \right) \tag{7.16}$$

Borne Inférieure de la capacité effective.

Afin de trouver des règles de dimensionnement faisant intervenir un nombre limité et facilement identifiable de paramètres, on fait l'hypothèse qu'il est possible de définir, pour chaque capacité «virtuelle», les paramètres suivants : valeur moyenne, valeur minimale et valeur maximale. Ces trois valeurs seront utilisées pour déterminer des bornes inférieures de la capacité effective. D'après la proposition précédente :

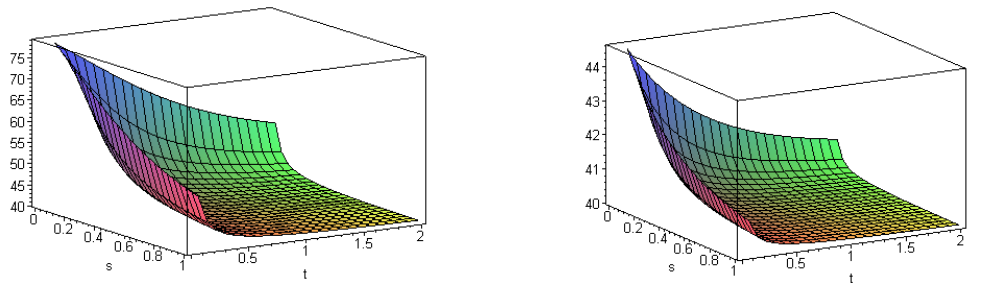
$$\begin{aligned} e^{-\theta x} &\leq \rho x + \gamma \\ \log \mathbb{E}[e^{-\theta x}] &\leq \log \left(\rho \mathbb{E}[x] + \gamma \right) \\ \frac{-1}{\theta t} \log \mathbb{E}[e^{-\theta x}] &\geq \frac{-1}{\theta t} \log \left(\rho \mathbb{E}[x] + \gamma \right) \end{aligned} \tag{7.17}$$

où ρ, γ vont dépendre de l'intervalle $[a, b]$ et de θ. Autrement dit, si l'on utilise la notation de capacités

effectives et si $c_{\min} t \leq c[0,t) \leq c_{\max} t$ et $\mathbb{E}[c[0,t)] = c_{\text{avg}} t$

$$\begin{aligned} \kappa(s,t) &\geq \frac{-1}{st} \log\left(\rho \mathbb{E}[c[0,t)] + \gamma\right) \\ \kappa(s,t) &\geq \frac{-1}{st} \log\left(e^{-sc_{\min} t} \frac{c_{\max} - c_{\text{avg}}}{c_{\max} - c_{\min}} + e^{-sc_{\max} t} \frac{c_{\text{avg}} - c_{\min}}{c_{\max} - c_{\min}}\right) \end{aligned} \tag{7.18}$$

La figure 7.4 illustre deux exemples de cette borne.



(a) $Cmin = 40, Cmax = 100, Cavg = 80$ (b) $Cmin = 40, Cmax = 100, Cavg = 45$

Fig. 7.4: Capacités Effectives : Bornes Linéaires

Remarque :

Avec une approche similaire à celle utilisée au paragraphe précédent, nous avons été amenés à considérer des bornes inférieures de la fonction exponentielle $x \rightarrow exp(sx)$ en considérant la famille des droites tangentes, telle qu'illustré sur la figure 7.3(b). Considérons l'intervalle $[x_1, x_2]$. Pour tout point $x_0 \in [x_1, x_2]$, nous avons :

$$\begin{aligned} e^{sx} &\geq e^{sx_0}\left(1 + s(x - x_0)\right) \Longrightarrow \\ \alpha(s,t) &\geq \lambda_0 + \log\left(1 + st(\lambda - \lambda_0)\right) \end{aligned}$$

où $\lambda = t^{-1}\mathbb{E}[x[0,t)]$ et $\lambda_0 \in [l,h]$; où $l = \operatorname{ess\,inf} x[0,t)/t$ est le débit minimal de la source et $h = \operatorname{ess\,sup} x[0,t)/t$ est son débit maximal (crête). Pourtant, la bande passante effective ainsi calculée ne vérifie pas certaines conditions nécessaires pour être considérée comme «une bande passante effective», notamment, parce qu'elle n'est pas croissante avec s.

7.3.1. Approximations Paraboliques

Les approximations précédentes, basées sur l'inégalité de Hoeffding, utilisent des approximations linéaires pour borner la famille de fonctions exponentielles et, ainsi, obtenir des bornes sur les bandes passantes (capacités) effectives. Une conséquence de cette approche est que les bandes passantes

effectives obtenues font intervenir les valeurs de l'intervalle considéré et les moyennes (premiers moments) des processus.

Nous avons étudié la possibilité d'utiliser des approximations paraboliques, avec l'idée de borner (inférieurement ou supérieurement) les fonctions exponentielles par une famille de paraboles, faisant intervenir également les moments de deuxième ordre.

Dans certains cas, ceci correspond à l'approximation de la fonction exponentielle autour d'un point par son développement de Taylor.

Approximation Parabolique de la fonction $e^{(sz)}$

$$\begin{aligned} y(z) &= a(z-z_0)^2 + b \\ y(z_a) &= a(z_a-z_0)^2 + b = e^{(sz_a)} \\ y(z_b) &= a(z_b-z_0)^2 + b = e^{(sz_b)} \end{aligned}$$

$$\begin{aligned} b &= e^{(sz_a)}\frac{(z_b-z_0)^2}{(z_b-z_0)^2-(z_a-z_0)^2} - e^{(sz_a)}\frac{(z_a-z_0)^2}{(z_b-z_0)^2-(z_a-z_0)^2} \\ a &= \frac{e^{(sz_b)}-e^{(sz_a)}}{(z_b-z_0)^2-(z_a-z_0)^2} \end{aligned}$$

Les équations précédentes définissent une famille de paraboles avec une erreur nulle aux extrémités de l'intervalle. La valeur de z_0 donne un troisième degré de liberté, et peut correspondre à : a) la parabole de pente 0 en $x_{\min}$ b) la parabole ayant la même pente en $x_{\min}$ que la fonction exponentielle ou c) des cas plus complexes faisant intervenir des optimisations pour trouver la parabole qui minimise une certaine distance, e.g.

$$\min_{z_0} \max_{z_a \le z \le z_b} \left| e^{(sz)} - y(z) \right|$$

Par exemple, si l'on cherche à déterminer une approximation pour une b.p.e., on peut considérer une parabole ayant son minimum en z_a (avec a > 0) ce qui donne une b.p.e. approximée par :

$$\hat{\alpha}(s,t) \approx \frac{1}{st}\log\left(\frac{e^{stx_{\max}}-e^{stx_{\min}}}{(x_{\max}-x_{\min})^2t^2}\left(\mathbb{E}[x[0,t)^2] - 2x_{\min}t\mathbb{E}[x[0,t)] + (x_{\min}t)^2\right) + e^{stx_{\min}}\right)$$

Il est a priori difficile de borner l'erreur commise. Afin d'avoir une idée de la validité de l'approximation nous illustrons cette approche dans un cas simple, dans lequel on suppose que le processus $X[0,t)$ suit une distribution uniforme en temps continu. Dans ce cas :

$$\begin{aligned} \alpha(s,t) &= \frac{1}{st}\log\left(\frac{e^{stx_{\max}}-e^{stx_{\min}}}{(x_{\max}-x_{\min})st}\right) \\ \mathbb{E}[X[0,t)] &= \frac{x_{\max}t + x_{\min}t}{2} \\ \mathrm{Var}[X[0,t)] &= \frac{({x_{\max}}^2 + {x_{\min}}^2 + x_{\max}x_{\min})t^2}{3} \end{aligned}$$

On peut donc comparer cette bande passante effective avec l'expression obtenue directement analy-

tiquement :

$$\begin{aligned}\mathbb{E}[e^{sX[0,t)}] &= \frac{1}{(x_{\max}-x_{\min})t}\int_{x_{\min}t}^{x_{\max}t} e^{sx}dx \\ &= \frac{e^{stx_{\max}}-e^{stx_{\min}}}{(x_{\max}-x_{\min})st} \\ \alpha(s,t) &= \frac{1}{st}\log\left(\frac{e^{stx_{\max}}-e^{stx_{\min}}}{(x_{\max}-x_{\min})st}\right)\end{aligned}$$

Remarquons que les approximations paraboliques peuvent également être obtenues pour les capacités effectives, de façon directe. Les approximations paraboliques peuvent être utiles lorsque l'on souhaite obtenir des approximations plus fines ou faire intervenir dans le modèle les moments de deuxième ordre. Par contre, il s'avère difficile de trouver des conditions pour définir des bornes inférieures. Par exemple, même avec des paraboles de pente zéro au point z_a, on ne peut pas assurer que l'expression obtenue soit une borne inférieure, car pour des valeurs de s (autrement dit, de θ) petites, la dérivée de la fonction exponentielle au voisinage de s est inférieure à celle de la parabole.

7.4. Formulation du Problème et Conditions d'Optimalité

Le problème du partage de charge à capacité variable peut être formulé de la façon suivante :

$$\min_{p_1,\ldots,p_L\in[0,1]} \mathbb{K} = \min_{p_1,\ldots,p_L\in[0,1]} \sum_{l=1..L} \mathrm{K}_l(p_l)$$

où

$$\begin{aligned}\mathrm{K}_l(p_l) &\triangleq p_l \exp(\mathrm{J}_l(p_l)) \\ \mathrm{J}_l(p_l) &\triangleq -N \inf_t I^R_{t,l}(b_l)\end{aligned}$$

avec les contraintes :

$$\begin{aligned}&\sum_{l=1..L} p_l = 1 \\ &0 \le p_l \le 1, \forall l \\ &\mathbb{E}[p_l x[0,t)] < \mathbb{E}[c_l[0,t)]\end{aligned}$$

Nous reprenons le problème d'optimisation proposé au chapitre précédent, à quelques détails près : le terme $(b+ct)s$ est remplacé par $bs + st\kappa(s,t)$

7.5. Exemples

Afin d'illustrer les principaux résultats obtenus, nous présentons ici deux exemples. Le premier est caractérisé par sa simplicité et le deuxième par son applicabilité directe. Les deux exemples seront présentés de la même façon : nous analysons d'abord une file d'attente isolée, afin de caractériser le point de travail et directement *la fonction de coût par route ou LSP.* Ensuite, nous illustrons la

fonction objective associée, *la fonction de coût du système* dans des cas à 2 LSPs. Finalement, nous présentons l'évolution du partage de charge optimale en fonction de certains paramètres importants.

Notons que dans tous les cas, il est nécessaire d'avoir une certaine connaissance des propriétés statistiques des processus de capacité. Nous insistons à ce sujet-là sur l'intérêt des études de la métrologie des canaux IP.

7.5.1. Exemple I : Processus Mouvement Fractionnaire Brownien

A titre d'exemple, motivé par le fait que les transformées de LogLaplace en sont analytiquement simples et dépendent d'un nombre réduit de paramètres, nous illustrons la démarche suivie dans le cas où les sources ainsi que les processus de service sont modélisées par des processus du type mouvement Brownien fractionnaire (fBm). La possibilité d'obtenir des expressions analytiques nous permet de mieux comprendre le rôle de chaque paramètre intervenant dans le modèle. Il est important de remarquer que l'utilisation de ce type de processus pose un certain nombre de problèmes, [58], entre autres le fait que d'après sa définition, un processus d'accroissements modélisé par un fBm peut avoir des incréments négatifs, situation impossible dans la réalité.

Notons que le modèle que nous avons présenté a été développé dans un contexte à temps discret, mais qu'il est facilement extensible à un contexte à temps continu (à certains détails techniques près). On utilise notamment pour cela le fait que les équivalents logarithmiques associés à l'asymptotique du grand nombre d'usagers ont aussi été démontrés dans un contexte à temps continu.

Caractérisation de la fonction de coût par LSP

D'une façon très similaire à celle utilisée dans le chapitre précédent, nous pouvons déterminer une expression simple du point de travail de chaque file, et donc de la fonction de coût par LSP associée. Considérons un LSP isolé, les transformées de LogLaplace associées aux sources et au processus de service sont (nous omettons l'index l par clarté) :

$$\Lambda^X_{t,p}(\theta) = \lambda p\theta t + \frac{1}{2}p^2\sigma_x^2\theta^2 t^{2H_x}$$
$$\Lambda^K_{t,p}(\theta) = M\theta t + \frac{1}{2}\Sigma_c^2\theta^2 t^{2H_c}$$
$$\Lambda^C_{t,p}(\theta) = \mu\theta t + \frac{1}{2}\sigma_c^2\theta^2 t^{2H_c}$$

Autrement dit, $M = N\mu$ et $\Sigma_c^2 = N^2\sigma_c^2$. Nous considérons par la suite $H_x = H_c$

$$I^R_t(b) = b\theta + (\mu - \lambda p)t\theta - \frac{1}{2}\left(\sigma_x^2 p^2 + \sigma_c^2\right)\theta^2 t^{2H}$$

Notons que, avec l'hypothèse de paramètres de Hurst égaux, le terme $I^R_t(b)$ correspond à celui d'un système à capacité fixe, servi par un serveur de capacité nominale égale à la capacité résiduelle moyenne $N\mu - N\lambda p > 0$ (par l'hypothèse de stabilité), alimenté par un agrégat de sources fBm où chaque source est de moyenne zéro, et de variance $\sigma_c^2 + p^2\sigma_x^2$. Autrement dit, le système équivaut à un système à capacité fixe égale à la capacité moyenne du système $N\mu$, alimenté par un agrégat de sources fBm avec une moyenne de l'agrégat de $Np\lambda$ et une variance égale à la somme des variances.

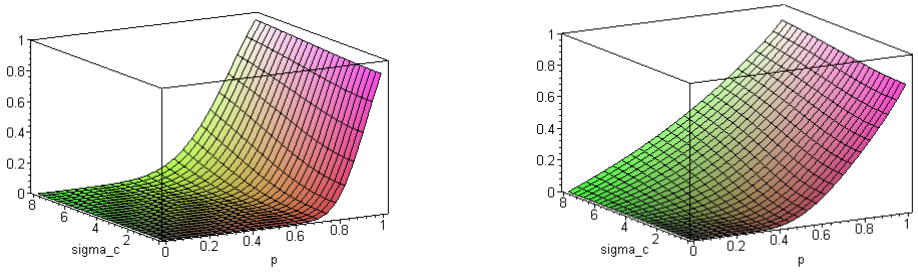
Le point de travail θ^* et le terme $I_t^R(b)|_{\theta^*}$ pour cette file d'attente isolée sont donnés par les expressions suivantes :

$$\theta^* = \frac{b + (\mu - \lambda p)t}{t^{2H}\left(p^2\sigma_x^2 + \sigma_c^2\right)}$$
$$I_t^R(b)|_{\theta^*} = \frac{1}{2}\frac{\left(b + (\mu - \lambda p)t\right)^2}{t^{2H}\left(p^2\sigma_x^2 + \sigma_c^2\right)}$$

Le point de travail t^* et la fonction de coût par LSP deviennent :

$$t^* = \frac{b}{\mu - \lambda p}\frac{H}{1 - H}$$
$$K(p) = p\exp\left(-N\frac{1}{2}\frac{\left(b + ((\mu - \lambda p)t^*\right)^2}{t^{*2H}(p^2\sigma_x^2 + \sigma_c^2)}\right)$$

La figure 7.5 montre l'évolution de la fonction de coût pour $N = 10, b = 1$ et $\lambda = 10$ en fonction de p et pour différentes valeurs de σ_c et la figure 7.6 pour différentes valeurs de μ

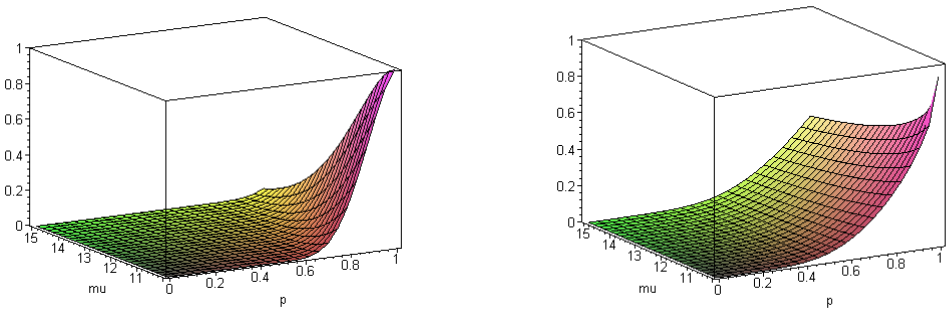


(a) $H = 0.8$ (b) $H = 0.2$

Fig. 7.5: Evolution des fonctions de coût par LSP, en fonction du partage p, $\lambda = 10$

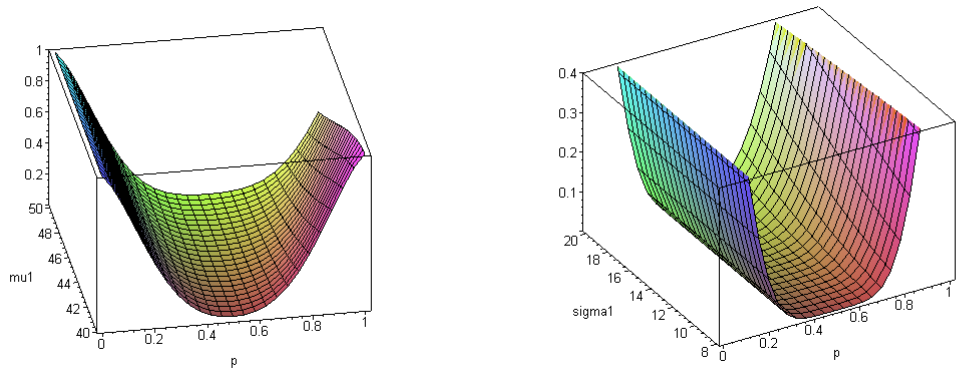
Partage de charge optimal

La figure 7.7 (a) illustre le partage de charge optimal pour deux LSPs, avec les paramètres suivants : $N = 10, \lambda = 40, \sigma_x = 20, H = 0.8$, premier LSP $\mu_1 \in [40.1..50], \sigma_1^c = 18$, deuxième LSP $\mu_2 = 40.1, \sigma_2^c = 18$. On peut remarquer son évolution : pour $\mu_1 = 40.1$, le système est symétrique et correspond à un partage optimal de $0.5, 0.5$. En augmentant la moyenne de la capacité du premier LSP, la fonction de coût favorise celui-ci. Dans (b), la fonction objective est représentée en fonction du partage p et de la variance associée au premier LSP. Notons qu'il est préférable d'avoir un système de capacité moyenne stable, car la fraction de trafic optimal allouée au premier LSP décroît avec σ_1^c. Notons également que dans l'exemple présenté ici, le système est stable pour toute valeur de p.



(a) $H = 0.8$ (b) $H = 0.2$

Fig. 7.6: Evolution des fonctions de coût par LSP, en fonction du partage p, $\lambda = 10$



(a) μ_1 (b) σ_1

Fig. 7.7: Partage de charge optimal, processus fBm.

Remarques

Si l'on augmente le nombre de sources multiplexées (avec $C_l = Nc_l$ et $B_l = Nb_l$ et les autres paramètres restant constants et en respectant les hypothèses de stabilité), on remarque l'effet du multiplexage statistique, car les fonctions de coût deviennent de plus en plus «plates» et n'augmentent qu'à très forte charge. Notons également qu'à faible et à moyenne charge, les fonctions de coût associées sont faibles, indiquant un «taux» de pertes réduit. Autrement dit, et de façon intuitive, à faible charge les différents partage de charge non optimaux sont néanmoins acceptables.

7.5.2. Exemple II : Borne Supérieure du Trafic et Processus de Service High-Low

On considère un système dans lequel N sources indépendantes constituent l'agrégat de trafic. Chaque source est caractérisée en utilisant la borne supérieure de sa bande passante effective (connaissant son débit moyen et son débit crête). Rappelons que la borne supérieure de la bande passante effective d'une source est donnée par :

$$\alpha(s,t) \leq \frac{1}{st}\log\left(\frac{x_{\max}-x_{\text{avg}}}{x_{\max}-x_{\min}}e^{sx_{\min}t}+\frac{x_{\text{avg}}-x_{\min}}{x_{\max}-x_{\min}}e^{sx_{\max}t}\right) \tag{7.19}$$

Dans le cas où $x_{\min}=0, x_{\max}\triangleq h$ et $x_{\text{avg}}\triangleq m$ l'expression précédente se simplifie et devient :

$$\alpha(s,t) \leq \frac{1}{st}\log\left(1-\frac{m}{h}+\frac{m}{h}e^{sht}\right) \tag{7.20}$$

Ce qui nous donne comme borne supérieure de la b.p.e. offerte par source :

$$\alpha_l(s,t) \leq \frac{1}{st}\log\left(1-\frac{m}{h}+\frac{m}{h}e^{sp_lht}\right) \tag{7.21}$$

L'objectif est d'évaluer l'effet des différents paramètres des processus de service et d'avoir des éléments de réponse à des questions comme «Quelle est l'erreur commise si l'on approxime un processus de service à deux états par un autre à capacité constante de même valeur que la moyenne du processus ?» ou «Quelle est l'évolution du partage de charge optimal, si l'on fait varier les moments des différents processus ?».

Caractérisation du processus de service

Les processus de service (un par file d'attente ou LSP) sont modélisés de la façon suivante : à chaque intervalle de temps, le système travaille à $K_{\max}$ avec probabilité p_{high} et à $K_{\min}$ avec probabilité p_{low}. Le service est supposé indépendant entre intervalles consécutifs. Nous définissons :

$$\begin{aligned} c_{\max} &\triangleq N^{-1}K_{\max} \\ c_{\min} &\triangleq N^{-1}K_{\min} \\ p_{high} &\triangleq \frac{\lambda}{\lambda+\mu} \\ p_{low} &\triangleq \frac{\mu}{\lambda+\mu} \end{aligned} \tag{7.22}$$

La transformée de LogLaplace et la capacité effective associées à ce processus sont données par les expressions suivantes :

$$\begin{aligned} \Lambda^C_{t,l}(\theta) &= \log\left(e^{\theta c_{max}t}\frac{\lambda}{\lambda+\mu}+e^{\theta c_{min}t}\frac{\mu}{\lambda+\mu}\right) \\ \kappa(s,t) &= \frac{-1}{st}\log\left(e^{-sc_{max}t}\frac{\lambda}{\lambda+\mu}+e^{-sc_{min}t}\frac{\mu}{\lambda+\mu}\right) \end{aligned} \tag{7.23}$$

Remarque :

Notons le rapport entre la capacité effective ainsi calculée et l'expression dérivée au paragraphe 7.3. En effet, la borne obtenue correspond au cas où :

$$\begin{aligned} p_{high} &\triangleq \frac{\lambda}{\lambda+\mu} = \frac{c_{\text{avg}} - c_{\min}}{c_{\max} - c_{\min}} \\ p_{low} &\triangleq \frac{\mu}{\lambda+\mu} = \frac{c_{\max} - c_{\text{avg}}}{c_{\max} - c_{\min}} \end{aligned} \tag{7.24}$$

$$\begin{aligned} \frac{\partial \Lambda_t^X(\theta,p)}{\partial\theta} &= \frac{mpht\theta e^{pht\theta}}{h-m+me^{pht\theta}} \\ \frac{\partial \Lambda_t^C(\theta)}{\partial\theta} &= \frac{\lambda e^{\theta c_{max}} c_{max} + \mu e^{\theta c_{min}} c_{min}}{\lambda e^{\theta c_{max}} + \mu e^{\theta c_{min}}} \end{aligned}$$

7.5.2.1. Système sans buffer

Nous ajoutons l'hypothèse «sans buffer» ($B_l = 0$), afin de supprimer les influences des tailles de buffers et de se centrer sur l'effet des capacités. Dans le contexte «sans buffer», rappelons que le point de travail temporel t* est égal à 1 (c'est-à-dire, que les débordements et pertes ont lieu sur un seul intervalle de temps, le rôle tampon du buffer n'existant plus). De plus, la complexité du problème est largement réduite. Dans ce cas, la valeur θ^* réalisant le supremum est solution de l'équation :

$$\theta_t^* = \operatorname{argsup}\left\{ b\theta - \log\left(e^{-\theta c_{max} t}\frac{\lambda}{\lambda+\mu} + e^{-\theta c_{min} t}\frac{\mu}{\lambda+\mu}\right) - \log\left(1 - \frac{m}{h} + \frac{m}{h}e^{\theta p_l h t}\right)\right\}$$
$$b = 0, t = 1$$

Dans notre cas,

$$\begin{aligned} \frac{\partial \Lambda_{1,l}^X(\theta,p)}{\partial\theta} &= -\frac{\partial \Lambda_{1,l}^C(-\theta)}{\partial\theta} \\ \frac{\lambda c_{max} e^{-\theta c_{max}} + \mu c_{min} e^{-\theta c_{min}}}{\lambda e^{-\theta c_{max}} + \mu e^{-\theta c_{min}}} &= \frac{mphe^{\theta hp}}{h-m+me^{\theta hp}} \end{aligned}$$

La figure 7.8 illustre les capacités effectives pour différentes valeurs de $\lambda, \mu, c_{\min}$ et $c_{\max}$, pour $t = 1$

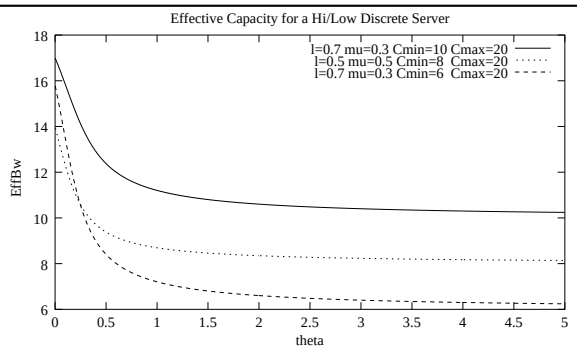
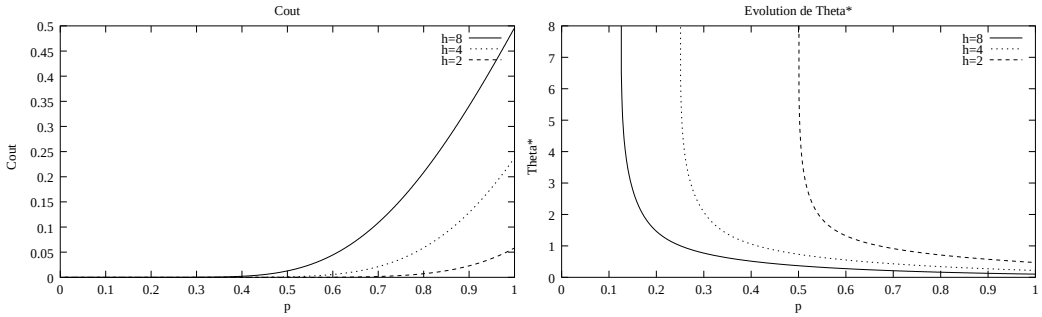


Fig. 7.8: Capacité Effective

Caractérisation de la fonction de coût par LSP

La figure 7.9 illustre (a) la fonction de coût par LSP et (b) l'évolution de la valeur de θ^* en fonction de la proportion p de trafic pour lequel on utilise la borne supérieure. Le trafic d'entrée est l'agrégation de $N = 10$ sources avec $m = 1.2$, pour différentes valeurs du débit crête (h). Le processus de service est défini par : $K_{\max} = 40, K_{\min} = 10, \lambda = p_{high} = 0.5$ et $\mu = p_{low} = 0.5$. La résolution de la valeur θ^* est réalisée numériquement à l'aide de l'équation précédente en utilisant une simple dichotomie avec condition d'arrêt $|\theta_a - \theta_b| \leq 0.001$.



(a) Coût LSP

(b) Valeur de *theta* optimale.

Fig. 7.9: Fonction de coût par LSP

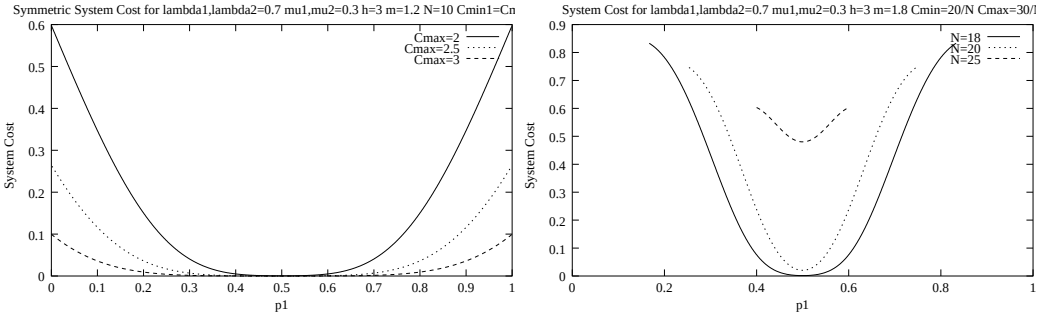
Caractérisation de la fonction de coût du Système

Dans cet exemple, nous considérons un système avec L=2 LSPs. La contrainte de partage de charge nous impose $p_2 = 1 - p_1$. Nous pouvons d'une façon grossière distinguer deux régimes de travail : à «faible charge» et à «forte charge». A faible charge (figure 7.10 (a)), le fonctions de coût présentent de zones plates et varient peu autour de l'optimum. Dans ce régime, de manière intuitive, des états non optimaux ne représentent (toujours relativement) presque aucune dégradation de performance. Au contraire, à forte charge, nous voyons que des contraintes de stabilité apparaissent explicitement, définissant des intervalles possibles et que les variations de la fonction de coût sont importantes dès que l'on s'éloigne de l'optimum. La figure (a) montre la fonction de coût pour une configuration particulière, dans laquelle les contraintes de stabilité sont inactives : $N = 10$, $m = 1.2$, $h = 3$, $\lambda_1 = \lambda_2 = 0.7$, $\mu_1 = \mu_2 = 0.3$, $K^1_{\min} = K^2_{\min} = 10$ et

- $K^1_{\max} = K^2_{\max} = 20 \qquad (c_{\max} = 2.0)$
- $K^1_{\max} = K^2_{\max} = 25 \qquad (c_{\max} = 2.5)$
- $K^1_{\max} = K^2_{\max} = 30 \qquad (c_{\max} = 3)$

Pour ces configurations, tout partage est possible, car on respecte toujours les contraintes de stabilité. Par contre, comme illustre la figure (b), si chaque source a $m = 1.8$, $h = 3$, et les processus de service sont déterminés par $\lambda_1 = \lambda_2 = 0.7$, $\mu_1 = \mu_2 = 0.3$, $K^1_{\min} = K^2_{\min} = 20$, $K^1_{\max} = K^2_{\max} = 30$, le nombre de sources multiplexées N définit l'ensemble admissible, $\forall l, Np_l m < \lambda K^l_{\max} + \mu K^l_{\max}$. Ainsi, les fonctions de coût du système (et l'ensemble admissible) sont tracés pour $N = 18, 20, 25$. En conséquence, plus le nombre de sources est grand, plus le coût du système est grand,

et plus l'ensemble des partages respectant la stabilité est petit. Evidemment, il est possible que pour certaines configurations, aucun partage ne respecte les contraintes de stabilité.



(a) Faible charge.

(b) Forte charge.

Fig. 7.10: Fonction de coût du système

Considérons un deuxième exemple. Cette fois, $N = 15$,$h = 3$,$m = 1.8$. Pour le processus de service 1, $\lambda_1 = 0.5$ et $\mu_1 = 0.5$, avec $K^1_{\min} = 15(c^1_{\min} = 1)$. Pour le processus de service 2, $\lambda_2 = 0.5$ et $\mu_2 = 0.5$, avec $K^2_{\min} = 15(c^1_{\min} = 1)$ et $K^2_{\max} = 45(c^2_{\max} = 3)$. La figure 7.11(a) montre la fonction de coût en fonction de p_1, pour différentes valeurs de $c^1_{\max}$. Le cas symétrique a un optimum de $(0.5, 0.5)$. Pour $c^1_{\max} = 1.1$, l'optimum $\mathbf{p} \approx (0.37, 0.63)$..

Afin d'évaluer l'impact des moments d'ordre supérieur, considérons encore un système avec $L = 2$, $N = 10$ sources i.i.d., $m = 1.7$, $h = 4$, $\lambda_1 = 0.5 = \lambda_2 = 0.5$, $\mu_1 = \mu_2 = 0.5$, $c^1_{\min} = 2$, $c^1_{\max} = 2$ (le premier LSP est vu comme un lien conservatif de capacité constante $K_1 = 20$)), avec :

- $c^2_{\min} = 0.0$, $c^2_{\max} = 4.0$.
- $c^2_{\min} = 1.0$, $c^2_{\max} = 3.0$.
- $c^2_{\min} = 1.5$, $c^2_{\max} = 2.5$.
- $c^2_{\min} = 2.0$, $c^2_{\max} = 2.0$.

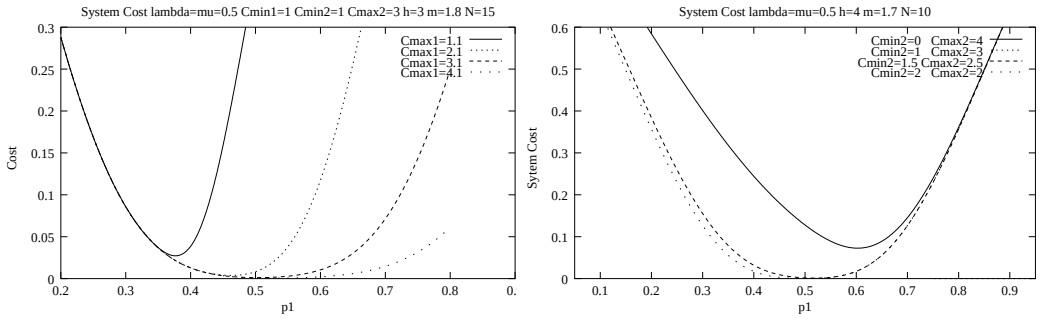
Comme illustré par la figure 7.11(b), plus l'intervalle de capacités est grand, à moyenne égale, plus la proportion de trafic allouée au LSP est petite. Nous reviendrons sur ceci dans la suite.

Partage de charge Optimal

La figure 7.12 montre l'évolution du partage de charge optimal pour deux files d'attente en fonction de la variation de la valeur maximale de la capacité du premier LSP (a) et par rapport à la variation de sa valeur minimale (b).

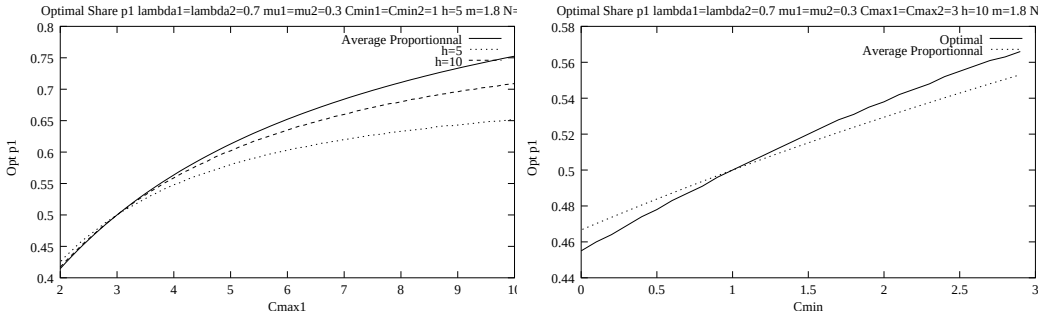
Insuffisance des approches basées sur des moyennes

Comme nous l'avons déjà montré dans les sections précédentes, la fonction de coût dépend fortement non seulement de la moyenne des processus de service mais aussi des autres moments. Pour illustrer l'insuffisance des approches basées sur un partage de charge proportionnel aux capacités moyennes, nous avons calculé le partage de charge optimal pour un système à deux LSPs, où la capacité moyenne



(a) $C_{\max 1}$ (b) $C_{\min 1}$

Fig. 7.11: Fonction Objective



(a) $C_{\max 1}$ (b) $C_{\min 1}$

Fig. 7.12: Evolution du Partage Optimal

de chaque LSP est la même. La figure 7.13 illustre cette notion : le partage de charge optimal dépend fortement de l'intervalle des valeurs, à moyenne constante (et est donc fortement lié à la variance des processus). Bien sûr, le partage optimal favorise le LSP de plus petite variance.

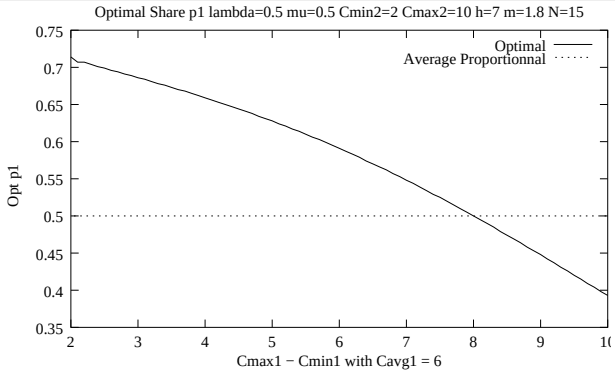


Fig. 7.13: Evolution du Partage Optimal

7.6. Un pas vers le partage de charge adaptatif...

7.6.1. Introduction

Les résultats précédents ont motivé l'étude de faisabilité d'une approche adaptative, pouvant présenter un certain intérêt vis-à-vis des changements des caractéristiques statistiques des processus intervenant dans les modèles. En effet, dans les systèmes réels, nous observons parfois une stationnarité locale, pendant un certain intervalle de temps, permettant l'application des mécanismes proposés. Néanmoins, des changements importants de la statistique des processus peuvent rendre inefficace le partage de charge calculé sous les hypothèses de stationnarité globale. Avec l'approche adaptative que nous présentons (ou une variante de celle-ci), le système peut s'adapter à ces changements.

7.6.2. Caractéristiques générales

1. Des conditions initiales définissent le partage de charge à appliquer.
2. A certains instants dits de *Mise à Jour* le système déclenche une procédure d'estimation des b.p.e. des sources et des capacités effectives des processus de service, en utilisant une fenêtre de temps connue et prédéterminée.
3. Une entité (par exemple le routeur d'entrée acheminant le trafic) calcule le nouveau partage de charge optimal, à partir des estimations des b.p.e. et le nouveau partage est déployé.

L'exemple le plus simple consiste en une mise à jour instantanée. A tout instant t, le système dispose d'une estimation des b.p.e. et le partage de charge à appliquer à l'instant $t+1$, noté $\mathbf{p}[t+1]$, est la solution du problème d'optimisation proposé dans la première partie du chapitre.

7.6.3. Modèle du Système

Le modèle que nous proposons est une variante du modèle présenté initialement. Il est illustré sur la figure 7.14. En résumé, à certains instants de *mise à jour* $t+1$, le routeur d'entrée calcule une estimation de la b.p.e. des sources $\hat{\alpha}(s,t)$ et des capacités effectives $\hat{\kappa}(s,t)$, en utilisant une fenêtre W (c.-à.-d. à partir des traces $(t-W,t]$), et le partage de charge à appliquer est solution du problème d'optimisation que nous avons présenté, en utilisant comme données d'entrée les estimations réalisées. Remarquons que du fait du caractère additif des b.p.e. des processus indépendants et de l'hypothèse que l'agrégat est constitué de sources i.i.d., il suffit de réaliser l'estimation *pour une seule source.* D'autre part, le mécanisme de partage de charge adaptatif nécessite de connaître, aux instants de mise à jour, la trajectoire récente des processus de service, hypothèse pouvant limiter l'applicabilité à des cas réels.

7.6.4. Simulations

Nous avons considéré deux estimations possibles : la première, l'estimateur de Dembo, et la deuxième en considérant les bornes de Hoeffding. Remarquons que le nombre limité de simulations que nous avons réalisées supposent que les arrivées dans deux intervalles de temps consécutifs sont indépendantes, hypothèse qui simplifie l'obtention de l'estimation.

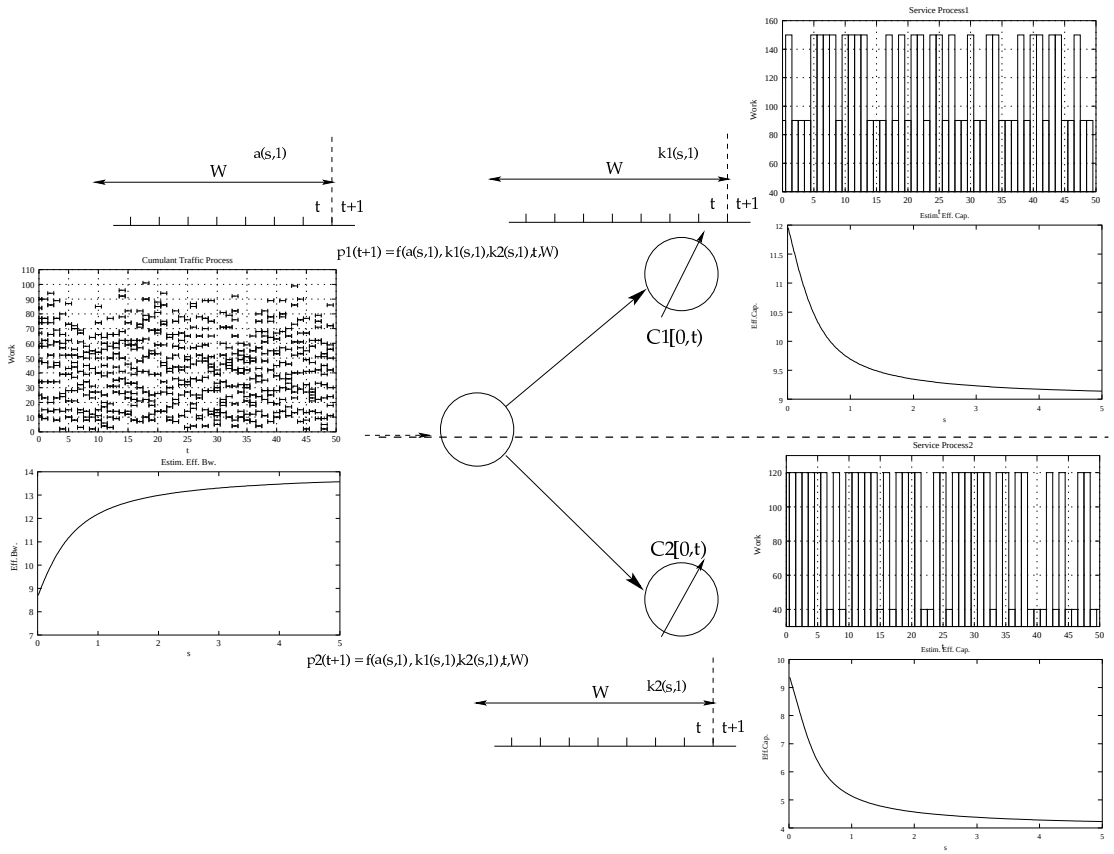


Fig. 7.14: Modèle du Système pour le Partage de charge Adaptatif

7.6.4.1. Estimateur de Dembo

Nous reprennons ici l'estimateur de Dembo présenté dans la section 5.7 pour l'estimation de bandes passantes et de capacités effectives.

Remarques

1. Avec les hypothèses du modèle $t^* = 1$, donc la taille des blocs est $b = 1$. Ceci facilite l'estimation des b.p.e.

2. Si $x[t_0]$ et $x[t_0 + k]$ sont indépendantes, on peut calculer les sommes X_i en utilisant des blocs de taille b espacés de k unités de temps.

3. Dans les simulations que nous présentons, le trafic d'entrée est composé de N sources, notées $x_i[t]$, de sorte que $x_i[t]$ et $x_i[t + 1]$ sont indépendantes, ce qui nous permet d'estimer les bandes passantes et les capacités effectives à partir de traces *consécutives*. Pour des modèles de trafic plus complexes (présentant une dépendance entre intervalles de temps consécutifs), l'échantillonnage de traces devrait prendre en compte la remarque 2.

Régimes Transitoires et Convergence

Une première simulation visait à avoir une idée de l'ordre de grandeur de la taille de la fenêtre nécessaire pour obtenir une estimation fiable. Il est difficile de donner des ordres de grandeur généraux, et la taille minimale de la fenêtre dépend des propriétés des sources et de leurs dépendances (par exemple, l'estimation des b.p.e. associées à des sources *classiques* est plus simple que celle de sources à mémoire longue). Nous avons considéré le cas suivant :

1. Le trafic d'entrée est un agrégat de $N = 10$ sources i.i.d.
2. Sur chaque intervalle de temps, le travail produit par une source $x_i[t] \sim U[8, 25]$
3. Le système est symétrique avec $L = 2$, et $C_l[t] \sim U[160, 180]$

Etant donnée la symétrie du problème, nous prévoyons un partage de charge optimal de $(0.5, 0.5)$. La figure 7.15 illustre l'évolution du régime transitoire.

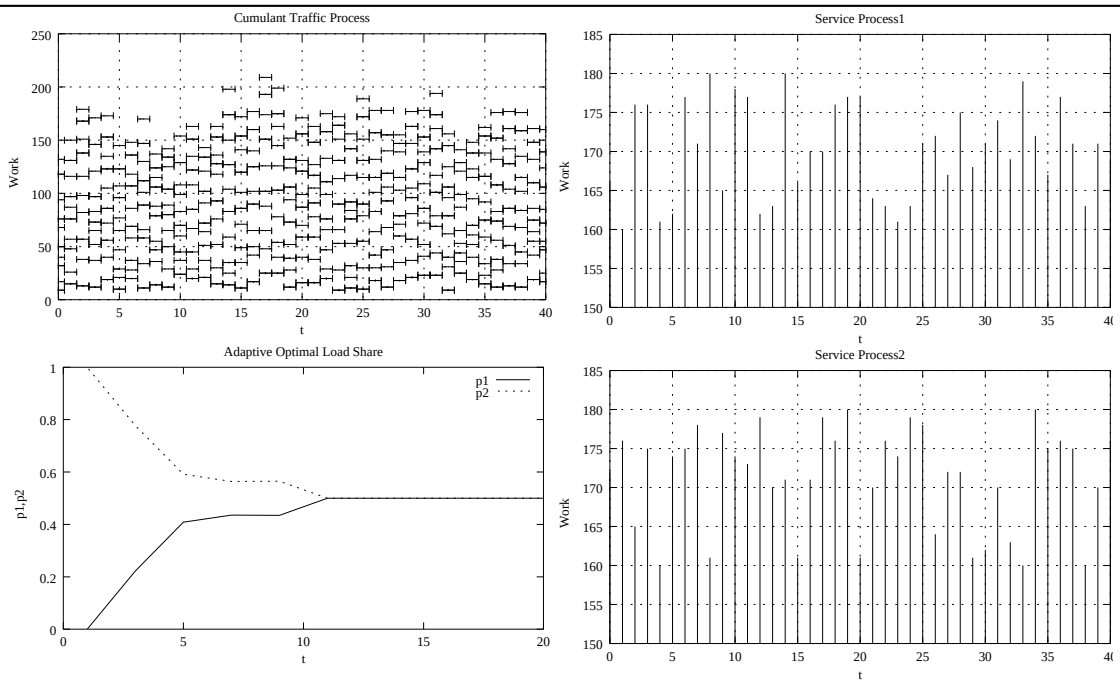


Fig. 7.15: Simulation et transitoires

Exemple

7.6.4.2. Partage Adaptatif avec les bornes de Hoeffding

L'estimation des b.p.e. est gourmande en temps de calcul. Néanmoins ce temps peut être réduit en considérant le même problème mais en estimant les b.p.e. et les capacités effectives en utilisant, respectivement, les bornes supérieures et inférieures données par les inégalités de Hoeffding.

La figure 7.17 montre l'évolution du partage de charge optimal avec un trafic d'entrée composé de $N = 10$ sources i.i.d. La quantité de travail produit par une source pendant un intervalle de temps

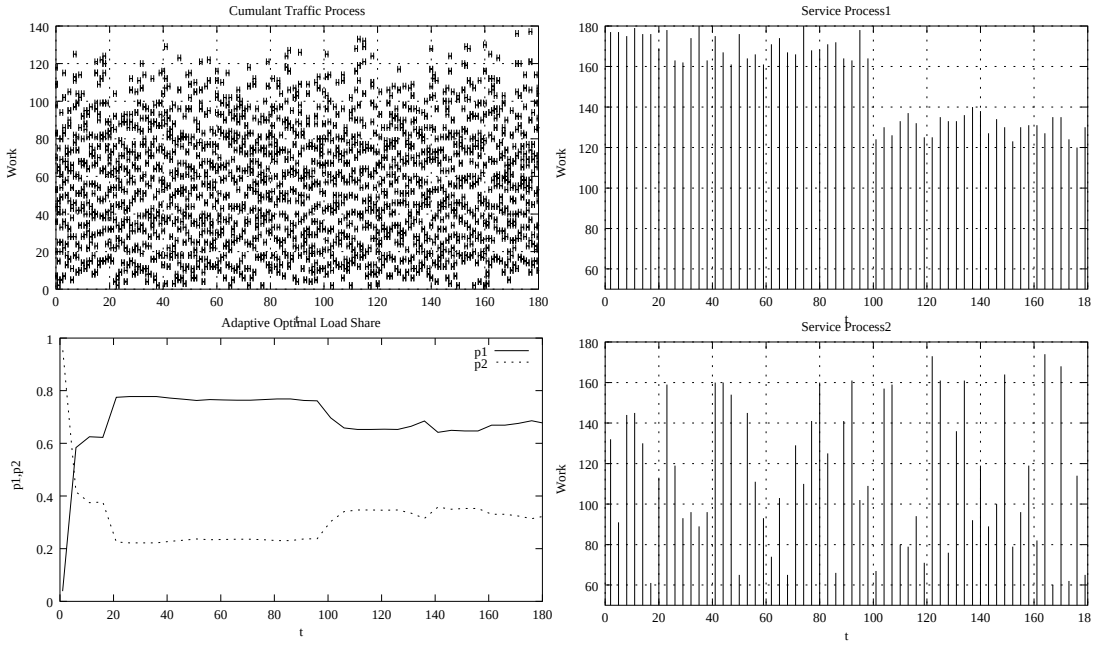


Fig. 7.16: Exemple I

correspond à une distribution discrète uniforme entre $[6, 26]$ unités de travail. Le système est composé de deux chemins. Le processus de service 1 présente 4 zones stationnaires :

1. $t \in [0, 100)$, travail $C_1[t] \sim U[160, 180]$
2. $t \in [100, 200)$, travail $C_1[t] \sim U[120, 140]$
3. $t \in [200, 300)$, travail $C_1[t] \sim U[60, 100]$
4. $t \in [300, 400)$, travail $C_1[t] \sim U[190, 220]$

Le processus de service 2 est globalement stationnaire avec une distribution $C_1[t] \sim U[60, 1800]$. Les simulations correspondent aux cas où la fenêtre de mesure W est a) 20 et b) 45 intervalles te temps.

Effet de la taille de la fenêtre de mesure

Dans les deux cas, le partage de charge optimal est capable de suivre les changements des propriétés stochastiques du système. Notons l'effet de la taille de la fenêtre W : une valeur relativement grande de la fenêtre a un effet lissant, mais il prend plus de temps à s'adapter à des changements stochastiques. Par contre, une valeur plus petite présente des oscillations légèrement plus marquées (dépendantes de la trajectoire), mais arrive à s'adapter plus rapidement aux différentes périodes stationnaires.

7.6.5. Remarques

1. *Intervalle Optimal :* le fait de considérer un système sans buffer implique que $t^* = 1$, ce qui nous permet d'estimer des b.p.e. pour cette valeur.

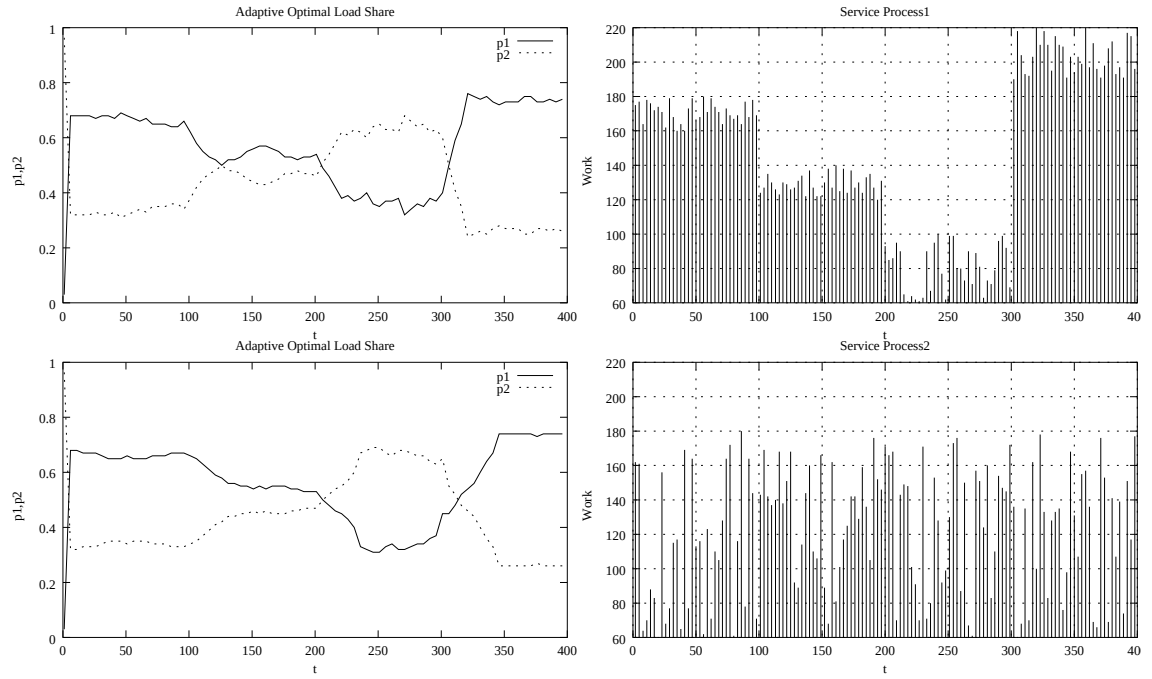


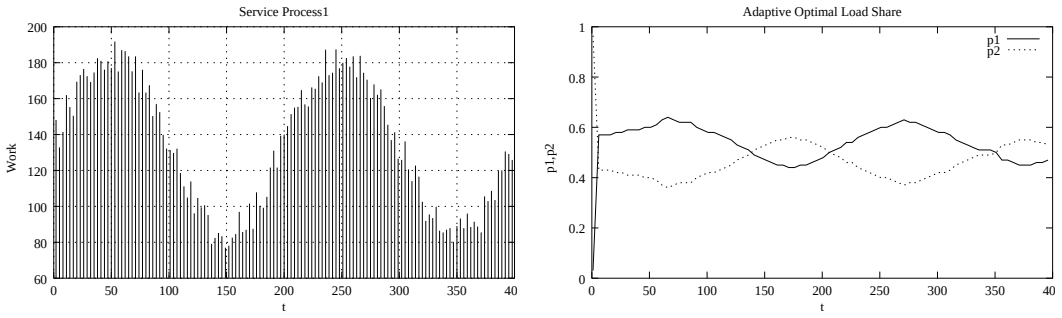
Fig. 7.17: Simulation avec estimation par bornes

2. *Indépendance :* les arrivées sur chaque intervalle sont indépendantes. Ceci nous permet de calculer un estimateur assez fiable (dont la précision dépend de la taille de la fenêtre de mesure). Pour des processus d'arrivées plus complexes, nous aurons besoin de méthodes d'estimation de b.p.e. plus efficaces, et notamment, d'obtenir des blocs de données indépendants.

3. *Oscillations :* cette approche présente des problèmes d'oscillation connus de longue date. En résumé : les périodes «stationnaires» doivent être suffisamment longues, et le calcul du partage de charge optimal aux instants de mise à jour doit pouvoir être réalisé en un temps relativement «petit» par rapport à la durée de l'intervalle de temps. Observons, par exemple, la figure 7.18 : le processus de capacité 1 est défini par : $x_i[n] = \max(U[100, 120] + 50\sin(4\pi n/T) + 25), 0)$, et le processus de capacité 2 est $C_2[t] \sim U[100, 120]$. L'algorithme présente des problèmes d'oscillation, dépendants entre autres facteurs de la taille de la fenêtre, car p_1 prend sa valeur maximale pour $t \approx 270$ et la sinusoïde prend sa valeur maximale pour $t \approx 250$.

4. *"Profiling" :* Nous nous sommes intéressés à la faisabilité de l'approche. L'application de techniques pour améliorer la performance du partage de charge adaptatif (amélioration des algorithmes, optimisation des calculs numériques, etc.) ainsi que la mise en place de mécanismes d'hystérésis restent des questions ouvertes.

5. *Utilisation de bornes :* Au vu de nos simulations, l'utilisation des bornes des b.p.e. et des capacités effectives donne des résultats largement satisfaisants, avec des temps de calcul très réduits en comparaison de l'approche utilisant l'estimation de Dembo.



(a) Trajectoire Capacité 1.

(b) Partage optimal.

Fig. 7.18: Partage Adaptatif : Problèmes d'oscillation

7.7. Extensions et Conclusions

7.7.1. Extensions à des Processus Markoviens

Le fait que pendant la réception de LSAs la capacité résiduelle est considérée constante peut amener à considérer les processus de capacité vus par le routeur ingress comme des processus Markoviens génériques, avec les équivalents des processus d'arrivées Markoviens *(Markov Arrival Process ou MAP, Markov Modulated Process)* pour les processus de service. Plusieurs auteurs ont obtenu des expressions pour les bandes passantes effectives associées mais l'exponentielle de matrices apparaissant dans les expressions nécessite des calculs matriciels complexes (développements en série, diagonalisation, puissances de matrices, entre autres). Une deuxième hypothèse consiste à limiter les processus de capacité à des processus de naissance et de mort (les sauts ne peuvent avoir lieu qu'entre deux états «consécutifs», hypothèse justifiée par le fait que les routeurs sont censés envoyer des LSA de mise a jour chaque fois qu'un certain seuil est dépassé). Cette deuxième hypothèse simplifie notablement les calculs, car les matrices deviennent quasi-diagonales.

Finalement, nous n'avons aucun élément pour déterminer la fonction de répartition de la variable aléatoire modélisant l'intervalle de temps entre la réception de deux LSA consécutifs (une première approche consistérait à considérer des v.a. exponentielles). Si l'on fait l'hypothèse que les LSA sont envoyés périodiquement, ceci correspondrait à une durée déterministe.

7.7.2. Conclusions

Dans ce chapitre nous avons proposé un modèle, similaire à celui du partage de charge sur la topologie multi-lien, pour l'évaluation du partage de charge dans un contexte de bout en bout, en utilisant la notion de capacité effective, pour des processus de service dont les caractéristiques statistiques sont connues. Encore une fois, l'idée de base est de modéliser la variabilité temporelle de la capacité associée à chaque LSP de façon simple. Nous soulignons le manque de procédures fiables pour l'obtention de telles statistiques, mais nous considérons que des travaux autour de la métrologie (techniques type «packet pair», etc.) peuvent s'avérer très utiles dans ce cadre.

A partir de la connaissance des propriétés stochastiques des processus de trafic et de service, nous proposons un problème d'optimisation non linéaire sous contraintes pour lequel nous avons obtenu des conditions d'optimalité (au moins locales) dont la complexité dépend quasi exclusivement des modèles des processus choisis. L'approche est illustrée à l'aide d'exemples qui ont été analysés en détail. Avec des modèles simples (Exemple II) on obtient déjà des résultats significatifs qui peuvent s'avérer suffisants.

Enfin nous avons proposé une solution adaptative, basée sur l'estimation des b.p.e. à partir de traces, nécessitant la connaissance des processus de service. Cette approche est cependant complexe et il est important de mieux comprendre les enjeux liés à l'estimation à partir de traces pour des cas non triviaux. Néanmoins, des solutions basées sur la mesure des débits crêtes, minimaux et moyens peuvent s'avérer suffisantes.

8. Extensions aux Réseaux

8.1. Introduction

Dans les chapitres précédents, nous avons étudié l'utilisation de l'asymptotique du grand nombre d'usagers dans le contexte de files d'attente isolées. Ce n'est que très récemment que la communauté scientifique a tenté d'étendre les expressions asymptotiques existantes à un contexte réseau de bout en bout. Wischik [106] a prouvé que la transformée de LogLaplace d'un seul flot composant de l'agrégat de trafic n'est pas modifiée du fait de traverser un nœud si ce flot est multiplexé avec un nombre important de flots similaires. Eun et Shroff [30] ont étudié un système de deux files d'attente en tandem, dans lequel la première file sert un nombre O(N) important de flots et un nombre fixe et fini de flots sont routés vers la deuxième. Leur conclusion est que la première file peut être ignorée en ce qui concerne le calcul de la probabilité de débordement quand N augmente. Néanmoins, leurs résultats ne s'appliquent pas quand le normbre de flots routés vers la deuxième file est aussi de l'ordre O(N).

D'une façon générale, on ne sait pas aujourd'hui caractériser le processus de sortie d'une file d'attente dans le contexte d'un agrégat de sources indépendantes. L'effet lissant d'un buffer de taille non négligeable rend cette tâche difficile, qui nécessite la connaissance de l'histoire du processus.

Néanmoins, certains résultats peuvent être étendus à des réseaux de files d'attente, notamment sous l'hypothèse de buffers de petite taille, plus précisement, si la taille du buffer croît moins vite que la dimension du système, c'est-à-dire :

$$\lim_{N\to\infty} \frac{B(N)}{N} \to 0 \tag{8.1}$$

Ozturk et al.[81] présentent les travaux récents autour des extensions aux réseaux de files d'attente et, sous l'hypothèse de petit-buffer, les auteurs ont obtenu des expressions asymptotiquement correctes pour la queue de distribution, appliquées aux réseaux de files d'attente en utilisant le principe de contraction et, en particulier, le résultat de la section 8.2. Les auteurs obtiennent également les régions d'admission associées.

Proposition 8.1.1 (Fonction de Taux de (Xn,Yn)). *Soient X_n, Y_n deux suites de variables aléatoires indépendantes, suivant des PGDs de fonctions de taux respectives $I_X(x)$ et $I_Y(y)$. La suite (X_n, Y_n) suit un PGD de fonction de taux $H_{x,y}(x,y) = I_X(x) + I_Y(y)$.*

Démonstration.

$$\begin{aligned}
\mathbb{P}\left(X_n \in \mathcal{B}, Y_n \in \mathcal{C}\right) &\overset{\text{indep}}{=} \mathbb{P}\left(X_n \in \mathcal{B}\right)\mathbb{P}\left(Y_n \in \mathcal{C}\right) \\
\frac{1}{n}\log \mathbb{P}\left(X_n \in \mathcal{B}, Y_n \in \mathcal{C}\right) &= \frac{1}{n}\log \mathbb{P}\left(X_n \in \mathcal{B}\right) + \frac{1}{n}\log \mathbb{P}\left(Y_n \in \mathcal{C}\right) \\
\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(X_n \in \mathcal{B}, Y_n \in \mathcal{C}\right) &= -\inf_{x\in\mathcal{B}} I_X(x) - \inf_{y\in\mathcal{C}} I_Y(y) \\
&= -\inf_{\substack{x\in\mathcal{B}\\ y\in\mathcal{C}}} \left\{I_X(x) + I_Y(y)\right\}
\end{aligned} \tag{8.2}$$

□

Théorème 8.1.1 (Principe de Contraction [23]). *Soient E, F deux espaces topologiques de Hansdorff. Soit X_n une suite de variables aléatoires à valeurs dans E qui suivent un PGD de fonction de taux $I(x)$. On fait l'hypothèse que les ensembles $\mathcal{X}_A \triangleq \{x : I(x) \leq A, \forall A > 0\}$ sont compacts. Soit $\phi : E \to F$ une fonction continue sur chaque ensemble $\mathcal{X}_A$. Alors, la suite $Y_n = \phi(X_n)$ à valeurs dans F vérifie un PGD de fonction de taux $J(y) = \inf_{x:\phi(x)=y} I(x)$.*

8.1.1. Multiplexage d'agrégats de trafic

Il est parfois intéressant de connaître les propriétés d'un agrégat de trafic et sa contribution à certains événements lorsqu'il est multiplexé sur une file d'attente. Considérons le système de la figure 8.1. L'agrégat $Z^{(N)}$ est la superposition des agrégats indépendants $X^{(N)}$ et $Y^{(N)}$. N correspond, comme toujours, au paramètre de dimension du système.

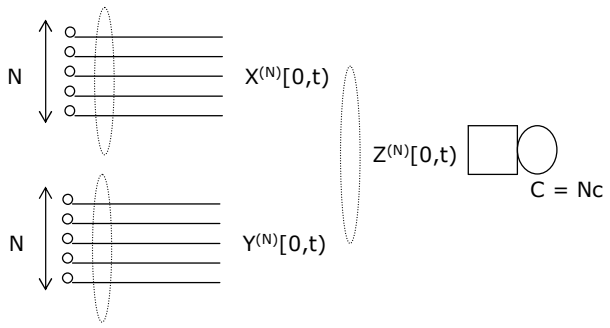


Fig. 8.1: Multiplexage d'agrégats de trafic

$$\begin{aligned}
Z^{(N)}[0,t) &= X^{(N)}[0,t) + Y^{(N)}[0,t) \\
I_t^Z(c) &= \sup_\theta \left\{\theta x - \frac{1}{N}\log \mathbb{E}[e^{Z[0,t)}]\right\} \\
I_t^Z(c) &= \sup_\theta \left\{\theta x - \frac{1}{N}\log \mathbb{E}[e^{X[0,t)+Y[0,t)}]\right\}
\end{aligned} \tag{8.3}$$

Proposition 8.1.2 (Fonction de Taux d'une somme (cf. [13])). *Soient X, Y deux variables aléatoires indépendantes de transformées de LogLaplace $\Lambda_X(\theta)$ et $\Lambda_Y(\theta)$. La transformée convexe de*

la transformée de LogLaplace de $Z = X + Y$ *notée* $\Lambda_Z^*(z)$ *vérifie :*

$$\begin{aligned}\Lambda_Z^*(z) &\triangleq \sup_\theta \{z\theta - \Lambda_Z(\theta)]\} \\ &= \inf_k \{\Lambda_X^*(k) + \Lambda_Y^*(z-k)\} \\ \Lambda_X^*(x) &\triangleq \sup_\theta \{x\theta - \Lambda_X(\theta)\} \\ \Lambda_Y^*(y) &\triangleq \sup_\theta \{y\theta - \Lambda_Y(\theta)\}\end{aligned} \tag{8.4}$$

La proposition précédente appliquée à notre cas nous permet d'affirmer que

$$I_t^Z(c) = \inf_{k>0} \left\{ I_t^X(k) + I_t^Y(c-k) \right\} \tag{8.5}$$

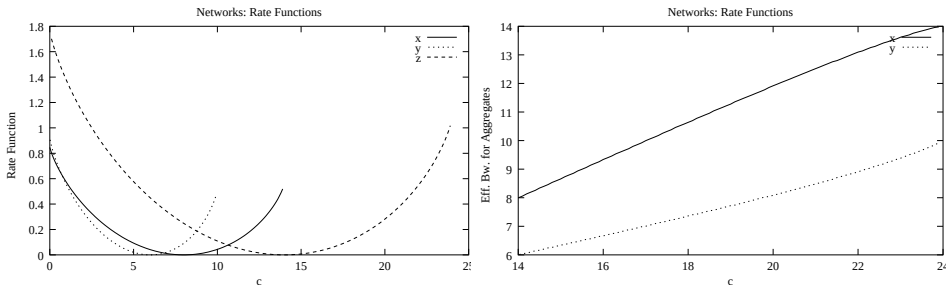
Sous l'hypothèse (8.1), les débordements ont lieu à $t = 1$. En notant :

$$I^X(x) \triangleq I_1^X(x) \qquad I^Y(y) \triangleq I_1^Y(y) \qquad I^Z(z) \triangleq I_1^Z(z) \tag{8.6}$$

on obtient :

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(W > B(N)) = -I^Z \tag{8.7}$$

La fonction de taux associée au PGD du travail cumulé dans la file d'attente peut être exprimée à l'aide des fonctions de taux des deux agrégats (8.5). Pour une capacité $C = Nc$, la valeur k^* qui réalise le minimum doit être interprétée comme la capacité attribuée à chaque source de l'agrégat X. La capacité résiduelle par source $c - k^*$ est donc attribuée à chaque source de l'agrégat Y.



(a) (b)

Fig. 8.2: Exemples numériques

La figure 8.2 en donne un exemple numérique. Les agrégats X et Y sont composés de sources pour lesquelles on utilise l'inégalité de Hoeffding pour borner inférieurement leur fonction de taux, avec comme débit crête et comme débit moyen respectivement $h_x = 14, m_x = 8, h_y = 10$ et $m_y = 6$. A gauche, (a) on représente les fonctions de taux respectives ainsi que la fonction de taux de l'agrégat Z. Notez que les fonctions de taux prennent leur valeur minimale aux débits moyens correspondants

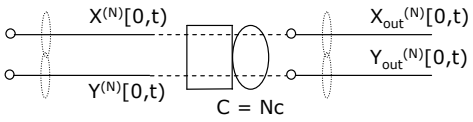
et sont finies sur l'intervalle $[0, h]$. A droite, (b), pour chaque valeur de c, les différents valeurs de k réalisant le minimum, valeur déterminant la capacité attribuée à chaque source des deux agrégats.

8.2. Caractérisation de l'agrégat de sortie

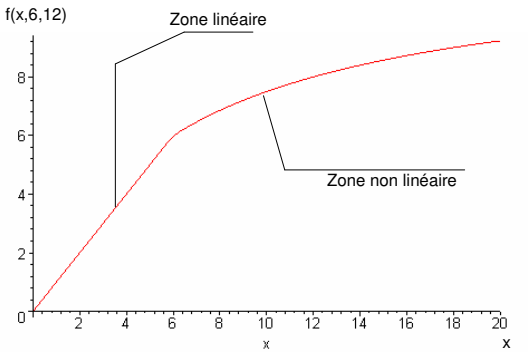
La caractérisation dans notre contexte de la fonction de taux associée à l'agrégat de sortie est due à Mazumdar et al., et utilise le principe de contraction à partir de la fonction de taux du même agrégat à l'entrée de la file d'attente.

Proposition 8.2.1 (Caractérisation de la fonction de taux de l'agrégat de sortie). *La fonction de taux de l'agrégat de sortie $X_{out}^{(N)}$ est donnée par l'expression :*

$$I^{X_{out}}(s) = \inf_{x,y|\frac{xc}{\max(x+y,c)}=s} \left\{ I^X(x) + I^Y(y) \right\} \tag{8.8}$$



(a)



(b)

Fig. 8.3: Fonction intervenant dans le Principe de Contraction pour la caractérisation de l'agrégat de sortie

$$f(x,y,z) = \frac{xz}{\max(x+y,z)} = \begin{cases} x & z > x+y \\ \frac{x}{x+y}z & z <= x+y \end{cases} \tag{8.9}$$

Interprétation de f(x,y,z)

La fonction $f(x, y, z)$ correspond à la notion assez intuitive de *bande passante allouée à l'agrégat x, en présence de y, avec une capacité z.* Pour des valeurs de y et z données, l'évolution de f(x,y,z) par rapport à x est illustrée figure 8.3. On peut y distinguer deux régions : la région linéaire et la région non linéaire. Dans la région linéaire, la file d'attente se comporte de façon transparente pour l'agrégat X qui nous intéresse. Dans la deuxième région, la quantité de trafic fluide que l'on obtient est proportionnelle à la capacité nominale du système et à la proportion x du trafic par rapport à l'agrégat total multiplexé (Z).

8.3. Analyse d'un réseau linéaire

Maintenant que nous sommes capables de déterminer la fonction de taux de sortie, analysons le réseau linéaire illustré figure 8.4 : l'agrégat étudié, $X^{(N)}$, traverse une suite de files d'attente. A chaque file d'attente i, l'agrégat est multiplexé avec un agrégat $Y_i^{(N)}$ *(cross traffic),* également composé d'un nombre N de sources indépendantes. On fait l'hypothèse que les agrégats $X^{(N)}$ et $Y_i^{(N)}$ sont indépendants et nous nous plaçons dans l'hypothèse de buffers de petite taille ($\lim_{N\to\infty} B(N)/N = 0$).

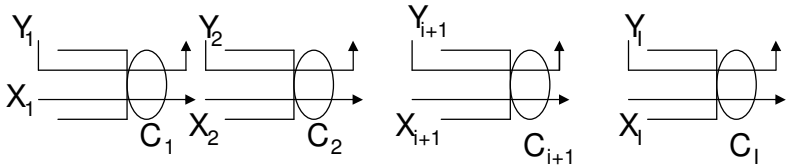


Fig. 8.4: Réseau Linéaire.

L'application directe des résultats précédents nous permet d'écrire le système d'équations suivant :

$$\begin{aligned} I_1(c_1) &= \inf_{y>0} \left\{ I^{X_1}(y) - I^{Y_1}(c_1 - y) \right\} \\ &\dots \\ I_i(c_i) &= \inf_{s>0} \left\{ I^{X_i}(s) + I^{Y_i}(c_i - s) \right\} \\ I_{i+1}(c_{i+1}) &= \inf_{s>0} \left\{ I^{X_{i+1}}(s) + I^{Y_{i+1}}(c_{i+1} - s) \right\} \\ &\dots \\ I_I(c_I) &= \inf_{y>0} \left\{ I^{X_I}(y) - I^{Y_I}(c_I - y) \right\} \end{aligned} \tag{8.10}$$

D'après (8.8),

$$I^{X_{i+1}}(s) = \inf_{x,y \mid \frac{x c_i}{\max(x+y, c_i)} = s} \left\{ I^{X_i}(x) + I^{Y_i}(y) \right\} \tag{8.11}$$

La complexité de ce système est principalement due à la difficulté de caractériser la fonction de taux associée à l'agrégat de sortie en fonction de son contexte de multiplexage (c.-à.-d. la capacité du serveur, la taille du buffer et l'agrégat de trafic externe) et de la fonction de taux associée à l'agrégat d'entrée. L'objectif de cette première étape est de déterminer les conditions qui nous permettent d'approximer les fonctions de taux de sortie par les fonctions d'entrée.

$$I^{X_i}(s) \approx I^{X_{i+1}}(s) \tag{8.12}$$

Soit ρ le débit moyen d'une source, et soit π son débit crête. D'après l'inégalité de Hoeffding,

$$\mathbb{E}[e^{\theta x(0,1)}] \leq \frac{\rho}{\pi} e^{\theta\pi} + \frac{\pi - \rho}{\pi} \tag{8.13}$$

Il est donc possible de déterminer une borne inférieure de la fonction de taux associée au processus (ici, avec l'hypothèse de N sources i.i.d.) :

$$\begin{aligned} I^X(x) &= \sup_\theta \left\{ \theta x - \frac{1}{N} \log \mathbb{E}[e^{\theta X(0,1)}] \right\} \\ I^X(x) &\geq \sup_\theta \left\{ \theta x - \log \left(\frac{\rho}{\pi} e^{\theta\pi} + \frac{\pi - \rho}{\pi} \right) \right\} \end{aligned} \tag{8.14}$$

La valeur θ^* réalisant le maximum est obtenue facilement car le terme $f(\theta, x, \pi, \rho)$ est concave par rapport à θ.

$$\begin{aligned} f(\theta, x, \pi, \rho) &= \theta x - \log \left(\frac{\rho}{\pi} \left(e^{\theta\pi} - 1 \right) + 1 \right) \\ \frac{\partial}{\partial\theta} f(\theta, x, \pi, \rho) &= x - \frac{\rho e^{\theta\pi}}{\frac{\rho}{\pi}\left(e^{\theta\pi} - 1\right) + 1} \\ \theta^* &= \frac{1}{\pi} \log \left(\frac{x(\pi - \rho)}{\rho(\pi - x)} \right) \end{aligned} \tag{8.15}$$

finalement

$$\begin{aligned} I^{X_i}(x) &= \frac{x}{\pi_{x_i}} \log \left(\frac{x}{\rho_{x_i}} \frac{(\pi_{x_i} - \rho_{x_i})}{(\pi_{x_i} - x)} \right) - \log \left(\frac{\pi_{x_i} - \rho_{x_i}}{\pi_{x_i} - x} \right) \\ I^{Y_i}(y) &= \frac{y}{\pi_{y_i}} \log \left(\frac{y}{\rho_{y_i}} \frac{(\pi_{y_i} - \rho_{y_i})}{(\pi_{y_i} - y)} \right) - \log \left(\frac{\pi_{y_i} - \rho_{y_i}}{\pi_{y_i} - y} \right) \end{aligned} \tag{8.16}$$

$$I^{X_{i+1}}(s) = \inf_{x,y \mid \frac{xc_i}{\max(x+y, c_i)} = s} \left\{ I^{X_i}(x) + I^{Y_i}(y) \right\} \tag{8.17}$$

Proposition 8.3.1 (Propriété de la transformée convexe (cf. Annexe A)). *Soit $\Lambda(\theta)$ une fonction croissante, convexe, telle que $\Lambda(0) = 0$. Sa transformée convexe, définie par $\Lambda^*(x) = \sup_\theta \{\theta x - \Lambda(\theta)\}$, prend sa valeur minimale en $x^* = \frac{\partial\Lambda(\theta)}{\partial\theta}|_{\theta=0}$.*

Proposition 8.3.2 (Invariance). *Soit $I^{X_{i+1}}(s)$ définie par (8.17),*

- *Si $\rho_{x_i} < s < c_i$ alors $I^{X_{i+1}}(s) \geq I^{X_i}(s)$.*
- *De plus, si $\rho_{x_i} < s < c_i - \rho_{y_i}$ alors $I^{X_{i+1}}(s) = I^{X_i}(s)$.*

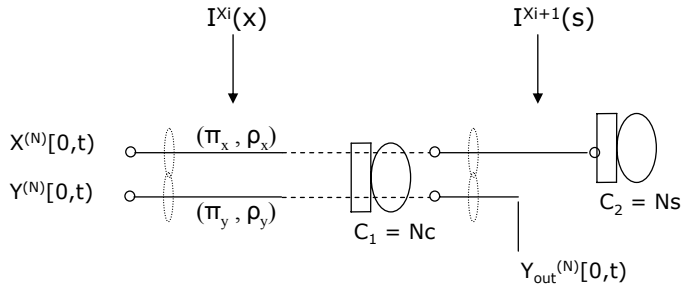


Fig. 8.5: Invariance

Démonstration.

$$I^{X_{i+1}}(s) = \inf\Big\{ \inf_{x,y\in I\!R | x=s, x+y\le c_i} \left\{I^{X_i}(x) + I^{Y_i}(y)\right\}, \inf_{x,y\in I\!R | \frac{xc_i}{x+y}=s, x+y>c_i} \left\{I^{X_i}(x) + I^{Y_i}(y)\right\} \Big\} \tag{8.18}$$

$$I^{X_{i+1}}(s) = \inf\Big\{ \inf_{y\in[0,h_y] | s+y\le c_i} \left\{I^{X_i}(s) + I^{Y_i}(y)\right\}, \inf_{y\in[0,h_y] | s+y>c_i} \left\{ I^{X_i}(\frac{s}{(c_i-s)y} + I^{Y_i}(y) \right\} \Big\} \tag{8.19}$$

Notons que pour $\rho_{x_i} < s < c_i$, $I^{X_i}(x)$ est croissante. En conséquence,

$$I^{X_i}(\frac{s}{c_i - s}y)\Big|_{s+y>c_i} \ge I^{X_i}(s) \tag{8.20}$$

et donc

$$I^{X_i}(\frac{s}{c_i - s}y) + I^{Y_i}(y) \ge I^{X_i}(s) + I^{Y_i}(y) \ge I^{X_i}(s) \tag{8.21}$$

car

$$I^{Y_i}(y) \ge 0 \qquad \forall y \in [0, h_y] \tag{8.22}$$

D'après la proposition (8.3.1), $I^{Y_i}(y)$ atteint son minimum en :

$$\begin{aligned} y^* &= N^{-1}\frac{\mathbb{E}[e^{\theta Y_i(0,1)}Y_i(0,1)]}{\mathbb{E}[e^{\theta Y_i(0,1)}]}|_{\theta=0} = N^{-1}\mathbb{E}[Y_i(0,1)] = \rho_{y_i} \\ I^{Y_i}(y^*) &= 0 \end{aligned} \tag{8.23}$$

finalement si $\rho_{x_i} < s < c - \rho_{y_i}$

$$0 = \inf_{y \in [0,h_y] | s+y \leq c_i} I^{Y_i}(y) \leq \inf_{y \in [0,h_y] | s+y > c_i} I^{Y_i}(y) \tag{8.24}$$

$$I^{X_{i+1}}(s) = I^{X_i}(s) + \inf_{y \in [0,h_y] | s+y \leq c_i} I^{Y_i}(y) = I^{X_i}(s) \tag{8.25}$$

□

Exemple

La figure 8.6 illustre les différentes fonctions de taux de sortie $I^{X_{i+1}}(s)$ d'un agrégat de trafic pour différentes valeurs de c_i. Dans l'exemple, les différents débits moyens et crêtes sont $\rho_{x_i} = 8, \pi_{x_i} = 14, \rho_{y_i} = 5.5, \pi_{y_i} = 10$. Les fonctions de taux de sortie peuvent être comparées à celle à l'entrée $I^{X_{i+1}}(s)$. Autre que la proposition précédente, on peut remarquer à quel point plus la capacité est élévée, plus la file présente un caractère transparent. Evidemment, pour des valeurs de $c_i \geq \pi_{x_i} + \pi_{y_i}$, la file i est complètement transparente.

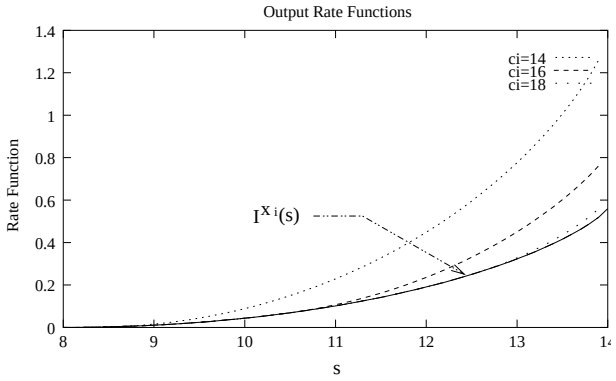


Fig. 8.6: Fonction de Taux de sortie d'un agrégat de trafic.

8.4. Interprétation et commentaires

Nous allons interpréter la proposition précédente à l'aide de la figure 8.5. De façon intuitive, sous l'hypothèse de stabilité, la valeur s peut être interprétée comme la capacité (normalisée par N) vue par l'agrégat X à l'entrée de la file $i+1$ *(downstream)*. Autrement dit, le calcul de métriques telles que la probabilité de pertes ou le taux de pertes de la file $i+1$ est réalisé en évaluant la fonction de taux $I^{X_{i+1}}(x)$ en s. Vis-à-vis de ce calcul, si la file i *(upstream)* a une capacité c_i telle que $s < c_i - \rho_{y_i}$, la file i a un comportement transparent, et on peut faire l'hypothèse que le passage par la file i ne modifie pas la fonction de taux de l'agrégat X. Ceci peut simplifier l'application à des réseaux de taille non triviale, car on n'est pas toujours obligé de calculer la fonction de taux à la sortie de la file i.

Perspectives

Les résultats présentés dans ce chapitre laissent la porte ouverte à de nouvelles études. A titre d'exemple, citons le routage dynamique de flots. Ceci serait réalisé en choisissant une route pour laquelle, au goulet d'étranglement, l'ajout d'un flot a un impact minimal sur une métrique telle que la probabilité de pertes ou le taux de pertes. Le calcul de cette métrique utiliserait les équations récursives présentées ici, en tirant si possible parti de la proposition d'invariance.

9. Ingénierie de trafic

Approche basée sur les bandes passantes effectives pour l'ingénierie de trafic dans les réseaux MPLS

9.1. Introduction et Motivation

L'extension aux réseaux des travaux présentés dans le présent document s'avère une tâche complexe. Néanmoins, avec l'objectif de trouver des règles applicables à un réseau, nous proposons dans ce chapitre une première approche pour l'ingénierie de trafic, et plus précisément, pour la répartition du trafic dans un réseau. L'approche présentée ici a un caractère exploratoire et est susceptible d'être améliorée lorsque l'on disposera de nouveaux éléments nous permettant de mieux comprendre le comportement asymptotique du trafic dans les réseaux, en particulier, comment dans le contexte de l'asymptotique du grand nombre d'usagers, le passage par un lien de transmission modifie les propriétés stochastiques des agrégats de trafic. L'approche que nous allons développer est caractérisée par les propriétés suivantes :

Heuristique

Nous allons formuler le problème de la répartition de trafic dans un réseau MPLS comme un problème d'optimisation globale pour lequel les fonctions de coût seront *basées* sur les résultats précédents, au lieu d'utiliser des fonctions de coût arbitraires. Nous utiliserons les transformées de LogLaplace des agrégats d'entrée et les fonctions de coût utiliseront l'asymptotique du grand nombre d'usagers qui nous permettra d'obtenir des équivalents logarithmiques des taux de pertes au niveau de chaque lien. Nous construirons une fonction objective du système pour laquelle les variables de décision correspondront à la répartition du trafic aux points d'entrée. Nous sommes conscients que d'un point de vue rigoureux, certaines des hypothèses que nous allons utiliser ne seront pas toujours vérifiées. Nous analysons ici ces hypothèses et leurs conséquences.

Contexte sans buffer

Un buffer de taille non négligeable (de l'ordre de O(N)) modifie les propriétés stochastiques des agrégats multiplexés d'une façon difficilement caractérisable, faisant intervenir l'histoire des processus concernés. Nous allons nous limiter aux cas où les buffers des liens de transmission du réseau sont de petite taille, condition que nous avons étudiée au chapitre 8. Cette hypothèse est cruciale : elle nous permet de quantifier l'effet de l'hypothèse d'invariabilité statistique, définie par la suite. En l'absence de buffers, l'effet lissant de la file d'attente est limité.

Invariabilité statistique

En utilisant les résultats développés au chapitre précédent, il est possible de caractériser les fonctions de taux des agrégats de sortie, sous l'hypothèse de buffers à croissance sous linéaire. Cependant, cette caractérisation devient trop complexe dans un réseau de taille moyenne, dont l'espace d'états croît de façon importante. Pour pallier à cette difficulté, nous allons faire l'hypothèse que dans ce contexte, la bande passante effective d'un flot est réduite, amincie *(ang. thinned)* par son passage à travers un lien, mais conserve ses *propriétés structurelles.* Cette caractérisation est analysée en détail dans la section 9.2.6.

Indépendance

Il est évident que, lors du passage par une file d'attente commune, les flots multiplexés ne sont à priori indépendants. Néanmoins, sous certains régimes asymptiques, les travaux de Wischik [106] ont donné certaines conditions concernant l'indépendance des flots de sortie, hypothèse nécessaire pour assurer l'additivité des bandes passantes effectives dans de topologies en tandem. En ce sens, nous allons faire l'hypothèse que les différents flots multiplexés sur un lien de transmission sont indépendants, afin de respecter le caractère additif des bandes passantes effectives : cette hypothèse est vraie seulement si les flots n'ont pas étés multiplexés auparavant.

9.2. Modélisation du système et notation utilisée

Comme toujours, N représente le paramètre du système intervenant dans l'asymptotique du grand nombre d'usagers. Chaque composante de la matrice de trafic (que nous allons caractériser formellement dans les sections suivantes) sera l'agrégat de N sources i.i.d. En conséquence, les capacités des liens de transmission seront du même ordre de grandeur O(N).

9.2.1. Le réseau vu comme un graphe...

- Soit $\mathcal{G} = (\mathcal{U}, \mathcal{L})$, le graphe orienté modélisant le réseau.
- Soit $\mathcal{U} = \{1, 2, \ldots, u, \ldots, U\}$, l'ensemble de nœuds du graphe, c'est-à-dire les LSRs du domaine MPLS. $\mathcal{U} = \mathcal{E} \cup \mathcal{P}$, ou $\mathcal{E}$ correspond à l'ensemble des nœuds de la périphérie du réseau (Edge LSR ou E-LSRs) et $\mathcal{P}$ est l'ensemble des LSRs du backbone. Nous allons définir des couples *Origine-Destination* $(a, b), a, b \in \mathcal{E}$.
- Soit $\mathcal{L} = \{1, 2, \ldots, l, \ldots, L\}$, l'ensemble des liens du graphe. Chaque lien $l \in \mathcal{L}$ est modélisé par une file d'attente sans buffer [1], de capacité totale $C_l = Nc_l$ où c_l est la capacité du lien par source. L'application de l'asymptotique du grand nombre d'usagers nous permettra d'obtenir, pour chaque lien, une approximation (un équivalent logarithmique) du taux de pertes, noté $LR(l) \in [0, 1]$. Nous définissons $q(l) = 1 - LR(l)$, avec $q(l) \in [0, 1]$. Ce terme correspond intuitivement au paramètre d'amincissement (réduction) des processus de trafic *(ang. thinning)* par le passage à travers le lien.
- Soit $\mathcal{OD} = \{1, 2, \ldots, g, \ldots, G\}$ l'ensemble des couples origine-destination. $g = (a, b)|a, b \in \mathcal{E}$. En accord avec la terminologie des chapitres précédents, nous appelons un couple origine-destination un *groupe de LSPs* ou simplement un *groupe.*

[1] Plus précisement, avec l'hypothèse des petits buffers, cf. chapitre précédent, $B(N)/N \to 0$

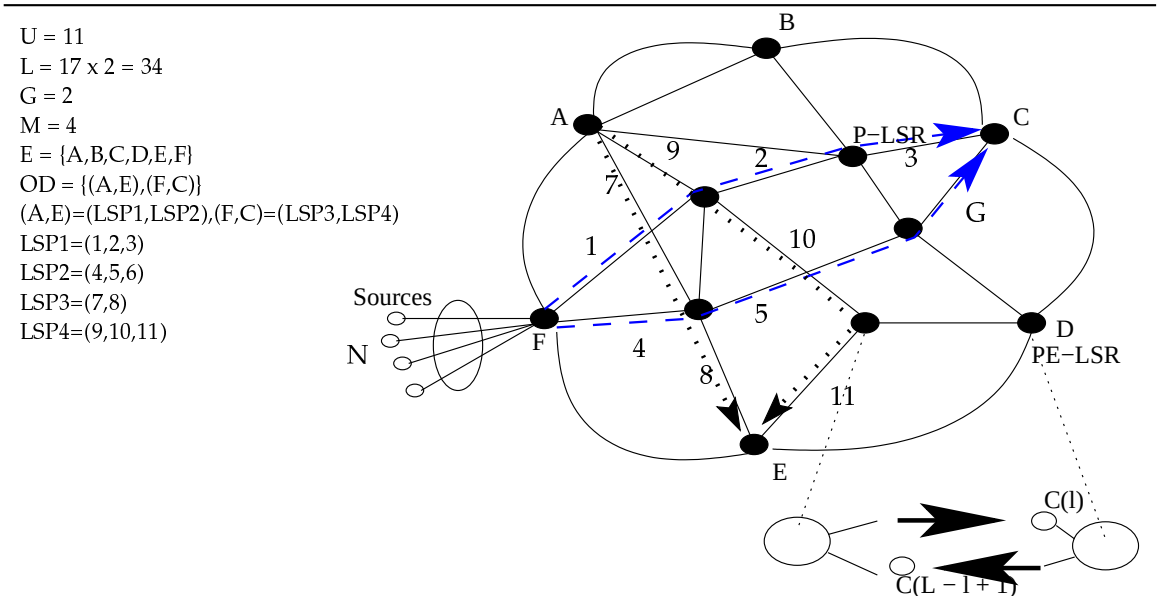


Fig. 9.1: Modèle du Système et notation utilisée

9.2.2. Calcul initial du placement (layout) des LSPs

– Soit $\mathcal{M} = \{1, 2, \ldots, m, \ldots, M\}$, l'ensemble des LSPs. Nous faisons l'hypothèse que les LSPs sont connus, et pré-établis. Nous verrons que dans le problème d'optimisation que nous allons proposer, la taille du vecteur des variables de décision est égale au nombre de LSPs dans le système (M). L'optimisation du placement des LSP pourrait être effectuée en considérant un nombre important de LSPs, et en rejetant les LSPs avec un partage négligeable.

– $\forall m \in \mathcal{M}, \forall g \in \mathcal{OD}$, on note $m \sim g$ si le LSP m appartient au groupe g. Autrement dit, si le groupe (couple) $g = (a, b), a, b \in \mathcal{E}$ alors a est le routeur d'entrée *(ang. Ingress LSR I-LSR)* et b est le routeur de sortie *(ang. Egress LSR)* du LSP m. Nous notons $\forall m \in \mathcal{M}$, $g(m) = g \in \mathcal{G} | m \sim g$.

9.2.3. Caractérisation du trafic d'entrée

– Pour chaque groupe $g \in \mathcal{OD}$, le trafic agrégat d'entrée (le trafic devant être acheminé entre le routeur d'entrée et le routeur de sortie définis par le groupe) est l'agrégation de N sources i.i.d. Chaque source est caractérisée par sa bande passante effective $\gamma_g(s, t)$. La b.p.e. de l'agrégat est donc $\Gamma(s, t) = N\gamma_g(s, t)$.

9.2.4. Caractérisation du trafic par LSP : la matrice de trafic effective

– Suivant l'approche utilisée dans cette thèse, la b.p.e. par source associée au trafic à acheminer sur chaque LSP d'un groupe donné $m \sim g$ est $p_m\gamma_g(p_m s, t)$, et sera notée $\alpha_m(s, t)$. Rappelons que $p \in [0, 1]$ est la proportion fluide de trafic par source.

– Pour tout LSP, $\forall m \in \mathcal{M}$, nous notons $h(m)$ sa longueur (en nombre de sauts) et nous notons sa route explicite $m = (l_m^1, l_m^2, \ldots, l_m^{h(m)})$ (liste des liens traversés par le LSP). Le terme $l_m^i \in \mathcal{L}$ représente le i-ème lien du LSP m. Nous notons également $l \sim m$ si le lien l appartient à m (le

LSP m passe par le lien l).

La figure 9.2 représente cette caractérisation.

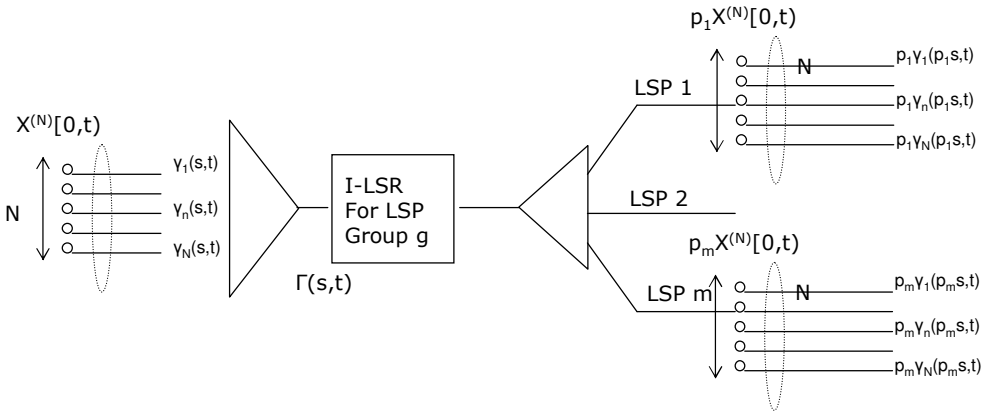


Fig. 9.2: Modèle du système et caractérisation du trafic par LSP, approche fluide : le paramètre p correspond à la proportion fluide de trafic à acheminer sur chaque LSP

La matrice de trafic effective est définie $\forall(a,b) \in \mathcal{E}x\mathcal{E}$ par :

$$MT = \begin{pmatrix} \gamma_{(1,1)}(s,t) & \gamma_{(1,2)}(s,t) & \cdots & \gamma_{1,E}(s,t) \\ \gamma_{(2,1)}(s,t) & \gamma_{(2,2)}(s,t) & \cdots & \gamma_{2,E}(s,t) \\ \vdots & \cdots & & \vdots \\ \vdots & \gamma_{(a,b)}(s,t)_{ij} & & \vdots \\ \vdots & \cdots & & \vdots \\ \gamma_{(E,1)}(s,t) & \gamma_{(E,2)}(s,t) & \cdots & \gamma_{E,E}(s,t) \end{pmatrix} \tag{9.1}$$

9.2.5. Caractérisation du trafic par lien

– Pour chaque lien $l \in \mathcal{L}$, chaque LSP le traversant ($\{m \in \mathcal{M} | l \sim m\}$) va lui *offrir* une b.p.e. par source notée $\alpha_{m,l}(s,t)$: b.p.e. par source offerte par le LSP m au lien l, et $\alpha_{m,l}(s,t) = 0$ si $l \not\sim m$. La b.p.e. totale par source offerte au lien l est notée $\alpha^{(l)}(s,t) = \sum_{\substack{m=1 \\ l\sim m}}^{M} \alpha_{m,l}(s,t)$.

– Ainsi, $\forall l$, $q(l)$ dépendra de N, de $\alpha_{m,l}(s,t)$ et de c_l.

9.2.6. Caractérisation du processus de sortie : Filtrage et amincissement

L'approche que nous suivons est la suivante : pour chaque lien, sous le régime asymptotique défini par l'hypothèse de grand nombre d'usagers, nous allons calculer une approximation (un équivalent logarithmique) du taux de pertes, $LR(l)$, dépendant du contexte de multiplexage, c'est-à-dire des

b.p.e. des flots concurrents et de la capacité du lien. L'hypothèse d'indépendance devient nécessaire : en faisant cette hypothèse, nous pouvons utiliser la propriété additive des b.p.e. associées.

La caractérisation exacte des processus de sortie, même sous l'hypothèse des petits buffers, nécessite l'application de techniques comme le principe de contraction (cf. chapitre 8). L'approximation que nous proposons dans le présent chapitre correspond à une transformation par *filtrage et amincissement spatial* de la bande passante effective : en effet, le terme $q(l) = 1 - LR(l)$ va déterminer un amincissement sur le paramètre spatial (s) de la bande passante effective : $T_l(\alpha) : \alpha_{in}(s,t) \rightarrow \alpha_{out}(s,t) = q(l)\alpha_{in}(q(l)s,t)$. Citons par exemple, le cas fBm, où $\alpha_{in}(s,t) = \lambda + \frac{1}{2}s\sigma^2 t^{2H-1}$, et $\alpha_{out}(s,t) = q(l)\lambda + \frac{1}{2}sq(l)^2\sigma^2 t^{2H-1}$: la moyenne du processus est amincie par $q(l)$ et la variance par $q(l)^2$. Intuitivement, avec un taux de pertes nul, (par exemple, les débits crêtes des sources plus petits que la capacité par source), le terme $q(l)$ vaut 1, la capacité devient «transparente» et la b.p.e. reste invariante. Par contre, avec un taux de pertes tendant vers 1, la b.p.e. de sortie tend vers zéro $\forall s, t$. Cette idée est illustrée par la figure 9.3 : la bandes passantes effectives illustrées correspondent à la borne supérieure en utilisant l'inégalité de Hoeffding :

$$\alpha_{in}^{i}(s,t) = \frac{1}{st}\log\left(1 - \frac{m_i}{h_i} + \frac{m_i}{h_i}e^{sth}\right)$$

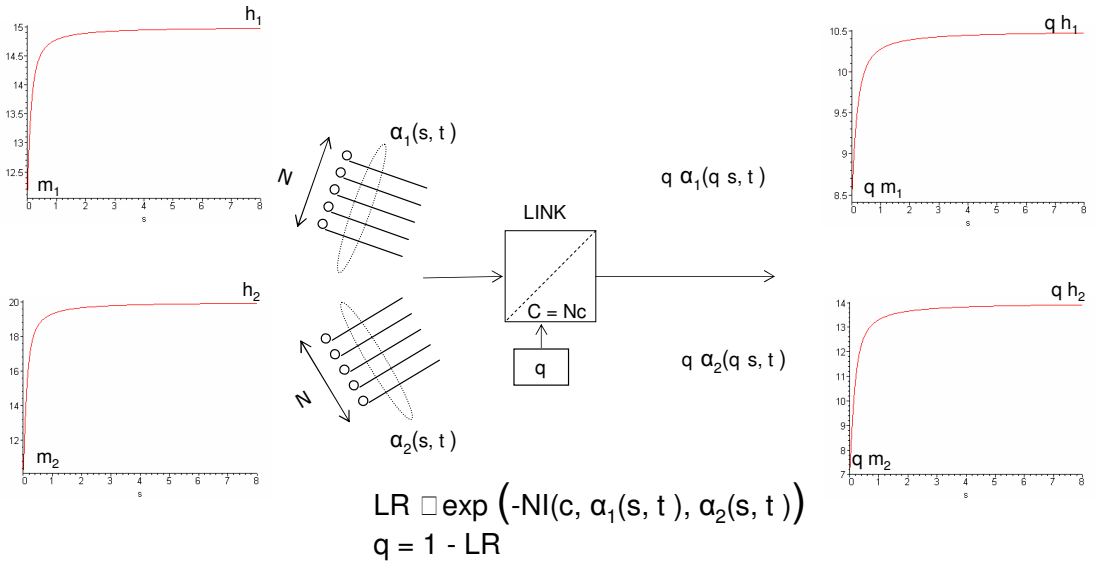


Fig. 9.3: Amincissement spatial, b.p.e. de Hoeffding q = 0.7, m1 = 12, h1=15, m2=10, h2=20

Notons l'analogie avec la caractérisation présentée dans la section 9.2.4.La complexité de notre approche vient du fait que le calcul de $q(l)$ fait intervenir les bandes passantes effectives *offertes,* elles-mêmes dépendantes des différents amincissements ayant eu lieu au niveau des liens précédant le lien considéré.

Discussion

- Un amincissement $\alpha(s,t) \rightarrow q\alpha(qs,t)$ où q est indépendant du processus d'arrivées est exact.
- Etant donné que le paramètre d'amincissement q est donné par le taux de pertes et que le processus de pertes et le processus d'arrivées ne sont pas indépendants, cette caractérisation est une approximation : intuitivement, plus le débit instantané est élevé, plus les pertes sont importantes.
- Si le nombre de LSPs (et donc de micro-flots) indépendants traversant un lien est suffisamment grand pour assurer qu'un processus d'arrivées (parmi tous ceux constituant l'agrégat total) est indépendant du processus de pertes (jusqu'à un certain degré), cette caractérisation peut être considérée comme acceptable.

Amincissement d'un LSP

Rappelons que nous avons noté un LSP $m \in \mathcal{M}$ comme $m = (l_m^1, l_m^2, \ldots, l_m^{h(m)})$. Nous définissons alors la fonction *index* $i_m(l)$, $\forall m \in \mathcal{M}, \forall l \in \mathcal{L}$ comme

$$i_m(l) = \begin{cases} j \quad | \quad l_m^j = l & \text{si } l \sim m \\ 0 & \text{sinon.} \end{cases}$$

c'est-à-dire que $i_m(l)$ renvoie l'index du lien l dans le vecteur que représente la route traversée par le LSP m.

La figure 9.4 montre l'amincissement suivi par le trafic acheminé sur un LSP lorsqu'il traverse un réseau de bout en bout.

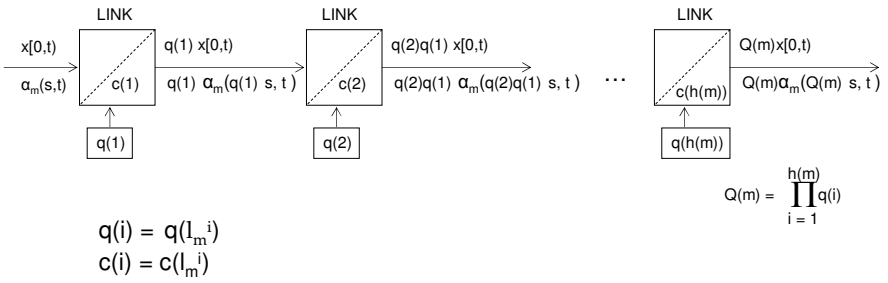


Fig. 9.4: Amincissement suivi par le trafic acheminé sur un LSP lors qu'il traverse un réseau de bout en bout.

Nous pouvons finalement écrire :

$$\begin{aligned} \alpha_{m,l}(s,t) &= \prod_{j=1}^{i_m(l)-1} q(l_m^j)\alpha_m \left(\prod_{j=1}^{i_m(l)-1} q(l_m^j)s,t \right) \\ \alpha^{(l)}(s,t) &= \sum_{\substack{m\in\mathcal{M} \\ l\sim m}} \alpha_{m,l}(s,t) \end{aligned} \tag{9.2}$$

La figure 9.5 utilise la représentation développée du graphe modélisant la topologie du réseau pour illustrer l'obtention de la b.p.e. offerte à un lien.

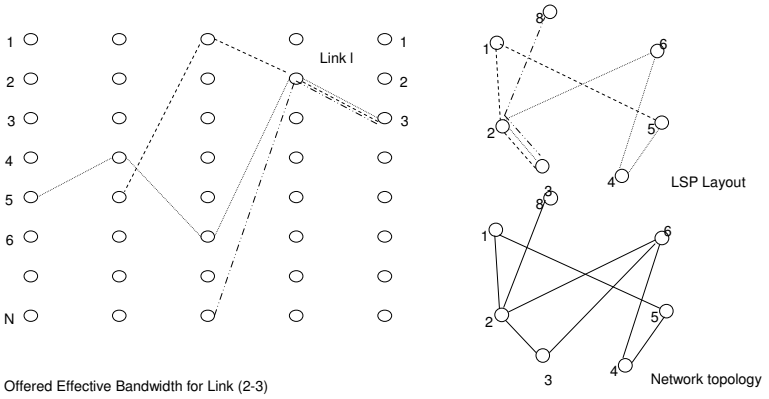


Fig. 9.5: Bande Passante Effective totale offerte au lien l, représentation en graphe développé.

Notation Matricielle

$$A_{ml} = \begin{pmatrix} \alpha_{1,1}(s,t) & \alpha_{1,2}(s,t) & \cdots & \alpha_{1,L}(s,t) \\ \alpha_{2,1}(s,t) & \alpha_{2,2}(s,t) & \cdots & \alpha_{2,L}(s,t) \\ \vdots & \cdots & & \vdots \\ \vdots & \alpha_{m,l}(s,t) & & \vdots \\ \vdots & \cdots & & \vdots \\ \alpha_{M,1}(s,t) & \alpha_{M,2}(s,t) & \cdots & \alpha_{M,L}(s,t) \end{pmatrix} \tag{9.3}$$

où

$$\alpha_{m,l}(s,t) = \begin{cases} \prod_{j=1}^{i_m(l)-1} q(l_m^j)\alpha_m \left(\prod_{j=1}^{i_m(l)-1} q(l_m^j)s,t \right) & \text{si} \quad l\sim m \\ 0 & \text{sinon.} \end{cases}$$

et

$$\alpha_m(s,t) = p_m \gamma_{g(m)}(p_m s,t) \qquad \text{avec} \gamma(s,t) \quad \text{connu.}$$

La b.p.e. par source associée au trafic offert au lien l peut s'écrire :

$$\alpha^{(l)}(s,t) = \mathbb{1}Ae_l = \sum_{\substack{m=1 \\ l\sim m}}^{M} \alpha_{m,l}(s,t) \tag{9.4}$$

où $\mathbb{1} = (1,1,\cdots,1)$ et $e_l = (0,\cdots,1,\cdots,0)^T$.

9.3. Formulation du problème et Répartition Optimale du trafic

Débit moyen offert par lien

$$\begin{aligned}\lambda^{(l)}(\mathbf{p}) &= \lim_{s\to 0}\alpha^{(l)}(s,t) = \lim_{s\to 0}\mathbb{1}Ae_l = \lim_{s\to 0}\sum_{\substack{m\in\mathcal{M}\\ l\sim m}}\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m\gamma_{g(m)}\left(\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m s,t\right)\\ &= \sum_{\substack{m\in\mathcal{M}\\ l\sim m}}\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m \lim_{s\to 0}\gamma_{g(m)}\left(\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m s,t\right) = \sum_{\substack{m\in\mathcal{M}\\ l\sim m}}\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m\lambda_{g(m)}\end{aligned} \tag{9.5}$$

car

$$\begin{aligned}\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m \lim_{s\to 0}\gamma_{g(m)}\left(\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m s,t\right) &= \lim_{s\to 0}(st)^{-1}\log\mathbb{E}\left[e^{s\prod_{j=1}^{i_m(l)-1} q(l_m^j)p_m X_{g(m)}(0,t]}\right]\\ &= \frac{\mathbb{E}\left[p_m\prod_{j=1}^{i_m(l)-1} q(l_m^j)X_{g(m)}(0,t]\right]}{t} = p_m\prod_{j=1}^{i_m(l)-1} q(l_m^j)\frac{\mathbb{E}[X_{g(m)}(0,t]]}{t}\\ &= p_m\prod_{j=1}^{i_m(l)-1} q(l_m^j)\lambda_{g(m)}\end{aligned} \tag{9.6}$$

En effet, d'après 9.5, le débit moyen par source offert au lien l est la somme des débits moyens par source des LSP le traversant, amincis par (a) le mécanisme de partage et (b) les différents taux de pertes des liens précédents. Nous avons considéré les deux critères suivants :

1. *Critère min-max,*

$$\min_{\mathbf{p}}\max_{l\in\mathcal{L}}\mathbb{P}_{\mathbf{p}}(W_l \geq B_l) \tag{9.7}$$

2. *Volume moyen de pertes*

$$\begin{aligned}\min_{\mathbf{p}}\sum_{l\in\mathcal{L}}\lambda^{(l)}(\mathbf{p})LR_{\mathbf{p}}(l) \approx \quad \min_{\mathbf{p}} &< \lambda(\mathbf{p}),\mathbf{e}(\mathbf{p}) >=\\ =\min_{\mathbf{p}}\sum_{l\in\mathcal{L}}\lambda^{(l)}(\mathbf{p})e^{NJ_l(\mathbf{p})}&\end{aligned} \tag{9.8}$$

avec

$$\begin{aligned} J_l(\mathbf{p}) &= -I_1(c_l, \alpha^{(l)}(s,t)) \\ &= -\sup_s \left\{ (c_l)s - s1\alpha^{(l)}(s,1) \right\} \end{aligned} \tag{9.9}$$

Remarquons que dans les deux cas nous faisons intervenir le même équivalent logarithmique.

9.3.1. Formalisation des contraintes

1. *Contrainte de Partage (I) :* Les variables de décision doivent correspondre au sens physique du partage de charge.

$$\forall g \in \mathcal{OD} \quad \sum_{\substack{m \in \mathcal{M} \\ m \sim g}} p_m = 1 \tag{9.10}$$

2. *Contrainte de Partage (II), ensemble faisable :* Les partages appartenant à un groupe doivent avoir pour somme 1. Autrement dit, tout trafic d'entrée doit être acheminé sur un des LSPs disponibles, avec une proportion qui dépendra de son partage.

$$\forall m = 1..M \quad p_m \in [0,1] \tag{9.11}$$

3. *Contrainte de Stabilité :* La contrainte de stabilité s'applique à tous les liens du graphe ; le débit moyen par source offert à un lien de transmission doit être inférieur à la capacité nominale par source du lien. Etant donné le caractère additif des b.p.e. ceci est équivalent à exiger que le débit moyen total offert à un lien de transmission ne dépasse pas sa capacité nominale. Avec la notation utilisée nous pouvons écrire :

$$\forall l \in \mathcal{L} \qquad \lambda^{(l)}(\mathbf{p}) < c(l) \tag{9.12}$$

Ainsi, nous avons $2M + L + G$ contraintes, (rappelons que M est le nombre de LSPs établis, L le nombre de liens orientés et G le nombre de couples origine-destination). Les contraintes sont linéaires et peuvent être formalisées de la façon suivante, avec leurs coefficients de Lagrange :

$$\begin{array}{ll} \text{Partage positif} & \\ c_m(p_m) = -p_m \leq 0 & u_1 \dots u_m \dots u_M \\ \text{Partage borné} & \\ c_m(p_m) = p_m - 1 \leq 0 & v_1 \dots v_m \dots v_M \\ \text{Partage} & \\ c_g(\mathbf{p}) = \sum_{\substack{m \in \mathcal{M} \\ m \sim g}} p_m - 1 = 0 & w_1 \dots w_g \dots w_G \\ \text{Stabilité} & \\ c_l(\mathbf{p}) = \lambda^{(l)}(\mathbf{p}) - c_l \leq 0 & z_1 \dots z_l \dots z_L \end{array} \tag{9.13}$$

Pour le critère du Volume Moyen de Pertes nous pouvons écrire le Lagrangien :

$$L(\mathbf{p},\mathbf{u},\mathbf{v},\mathbf{w},\mathbf{z}) = f(\mathbf{p}) - \sum_{m=1}^{M} u_m p_m + \sum_{m=1}^{M} v_m(1-p_m) + \sum_{g=1}^{G} w_g \left(\sum_{\substack{m\in\mathcal{M}\\ m\sim g}} p_m - 1 \right) + \sum_{l\in\mathcal{L}} z_l \left(\lambda^{(l)}(\mathbf{p}) - c_l \right) \tag{9.14}$$

avec

$$f(\mathbf{p}) = \sum_{l\in\mathcal{L}} \lambda^{(l)}(\mathbf{p}) e^{NJ_l(\mathbf{p})} \tag{9.15}$$

Les dérivées partielles correspondantes s'écrivent

$$\begin{aligned} \frac{\partial L(.)}{\partial p_m} &= \frac{\partial f(\mathbf{p})}{\partial p_m} - u_m - v_m + \sum_{g=1}^{G} w_g \mathbb{1}_{\{m\sim g\}} + \sum_{l\in\mathcal{L}} z_l \frac{\partial \lambda^{(l)}(\mathbf{p})}{\partial p_m} \frac{\partial f(\mathbf{p})}{\partial p_m} \\ &= \sum_{l\in\mathcal{L}} \left\{ \frac{\partial \lambda^{(l)}(\mathbf{p})}{\partial p_m} e^{NJ_l(\mathbf{p})} + \lambda^{(l)}(\mathbf{p}) e^{NJ_l(\mathbf{p})} N \frac{\partial J_l(\mathbf{p})}{\partial p_m} \right\} \\ &= \sum_{l\in\mathcal{L}} e^{NJ_l(\mathbf{p})} \left\{ \frac{\partial \lambda^{(l)}(\mathbf{p})}{\partial p_m} + \lambda^{(l)}(\mathbf{p}) N \frac{\partial J_l(\mathbf{p})}{\partial p_m} \right\} \end{aligned} \tag{9.16}$$

$$\begin{aligned} &\frac{\partial L}{\partial u_m} = -p_m; \qquad \frac{\partial L}{\partial v_m} = p_m; \qquad \frac{\partial L}{\partial w_g} = \sum_{\substack{m=1\\ m\in g}}^{M} p_m - 1 \\ &\frac{\partial L}{\partial z_l} = \lambda^{(l)}(\mathbf{p}) - c_l = r_l(\mathbf{p}) \quad \text{capacité résiduelle moyenne.} \end{aligned} \tag{9.17}$$

9.4. A propos de la complexité...

«Profiling»

L'heuristique proposée repose essentiellement sur le calcul de l'équivalent logarithmique du taux de pertes, approximé par une fonction exponentielle avec une fonction de taux. Même si pour certains modèles de trafic la fonction de taux peut être calculée explicitement, ceci n'est pas le cas pour des bandes passantes effectives arbitraires. La méthode fait appel en général à une résolution numérique.

Le taux de pertes

Nous avons utilisé pour le taux de pertes son équivalent logarithmique. Sans pénalisation de complexité importante, nous pouvons aussi utiliser les expressions asymptotiquement exactes [58]. Les auteurs ont montré que négliger le terme devant l'exponentielle est une approche conservative, justifiant a priori notre approche. Cependant, le taux de pertes définit aussi l'amincissement suivi par les différents flots (autrement dit, les b.p.e. de sortie seront plus petites que les b.p.e. réelles).

9.5. Exemples

Nous illustrons ici deux topologies différentes et les effets des paramètres sur la fonction de coût. Dans les deux cas, nous modélisons le trafic en utilisant la borne supérieure de sa b.p.e. ce qui nécessite la connaissance du débit crête et du débit moyen. Si les composantes de la matrice de trafic utilisent la borne de Hoeffding (ρ est le débit moyen et π le débit crête), nous avons :

$$\begin{aligned} \gamma_g(s,t) &= \frac{1}{st}\log\left(1-\frac{\rho_g}{\pi_g}+\frac{\rho_g}{\pi_g}e^{s\pi_g t}\right) \\ \alpha_m(s,t) &= \frac{1}{st}\log\left(1-\frac{\rho_{g(m)}}{\pi_{g(m)}}+\frac{\rho_{g(m)}}{\pi_{g(m)}}e^{p_m s\pi_{g(m)} t}\right) \qquad \forall m\sim g \end{aligned} \tag{9.18}$$

Au niveau de chaque file d'attente, avec l'hypothèse buffer-less, les pertes sont instantanées, et le point de travail $t^*=1$. La fonction de taux devient :

$$I_l = \sup_s\left\{(c_l)s - \sum_{\substack{m\in\mathcal{M}\\ l\sim m}}\log\left(1-\frac{\rho_{g(m)}}{\pi_{g(m)}}+\frac{\rho_{g(m)}}{\pi_{g(m)}}e^{p_m(\prod_{j=1}^{i_m(l)-1}q(l_m^j))s\pi_{g(m)}}\right)\right\} \tag{9.19}$$

Pour le calcul de la fonction de taux du lien l, le paramètre s^* (point de travail spatial) est solution de :

$$c_l = \sum_{\substack{m\in\mathcal{M}\\ l\sim m}} \frac{p_m(\prod_{j=1}^{i_m(l)-1}q(l_m^j))\pi_{g(m)}}{1-\frac{\rho_{g(m)}}{\pi_{g(m)}}+\frac{\rho_{g(m)}}{\pi_{g(m)}}e^{p_m(\prod_{j=1}^{i_m(l)-1}q(l_m^j))s\pi_{g(m)}}} \tag{9.20}$$

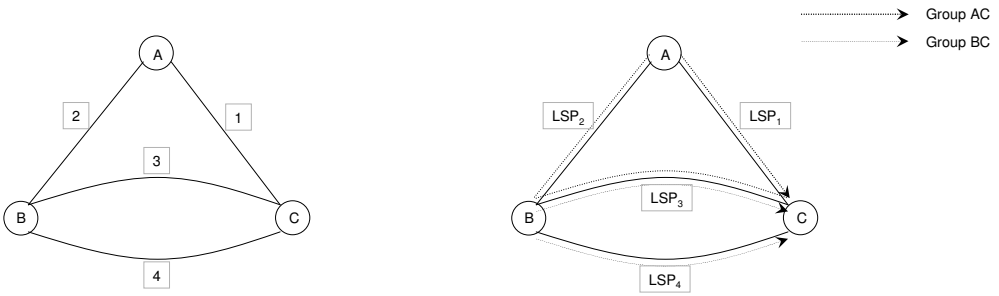
Le calcul du point de travail et de la fonction de taux avec les b.p.e. de Hoeffding peut être réalisé en utilisant les expressions analytiques (cf. chapitre précédent) si les groupes de LSPs sont disjoints (un seul agrégat de sources est injecté sur un lien donné). Avec un degré maximum de 2 (c.à.d, au plus 2 LSPs différents traversent un lien donné), nous pouvons aussi utiliser l'expression de l'équation (8.4). Par la suite, nous allons considérer principalement le critère *volume moyen de pertes,* sauf mention explicite.

9.5.1. Exemple I

La première topologie analysée est illustrée par la figure 9.6. La caractéristique la plus notable de cette topologie est que le lien BC1 est partagé par les LSP 2 et 3. Nous avons 4 LSPS candidats, regroupés en deux groupes.

Effet du Multiplexage Statistique (N)

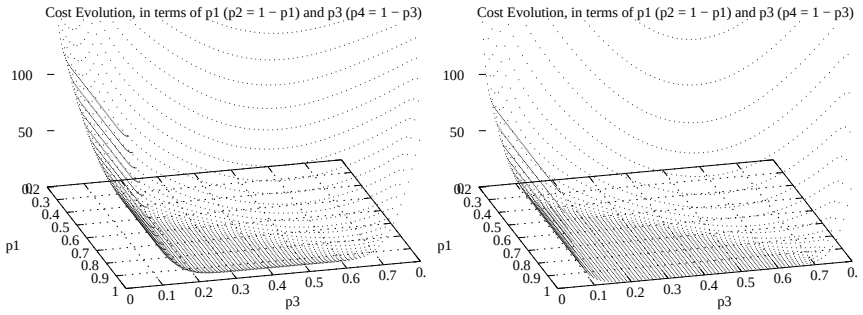
La première instance du problème est la suivante : Chaque agrégat est superposition de $N=50$ sources caractérisées par leur b.p.e. de Hoeffding. Chacune de ces sources a pour coefficients $\rho=8$ et $\pi=20$. Les capacités des liens de transmission sont $C_{AC}=N12, C_{AB}=N10, C_{BC1}=N8$ et $C_{BC2}=N10$. La figure 9.7 répresente la fonction de coût du système pour (a) N = 50 et (b) N = 100.



(a) (b)

Fig. 9.6: Exemple I : topologie et placement des LSPs. B.P.E utilisant la borne supérieure de Hoeffding.

Dans les deux cas, l'optimum est atteint en $p_1 = 0.77, p_2 = 0.23, p_3 = 0.36, p_4 = 0.64$. Etant donné que les capacités sont de la forme $C_l = Nk_l$, on peut observer l'effet du multiplexage statistique : la fonction de coût dans la figure (b) présente un caractère plus flat autour de l'optimum.



(a) $N = 50$ (b) $N = 100$

Fig. 9.7: Exemple I : effet du Multiplexage Statistique. N

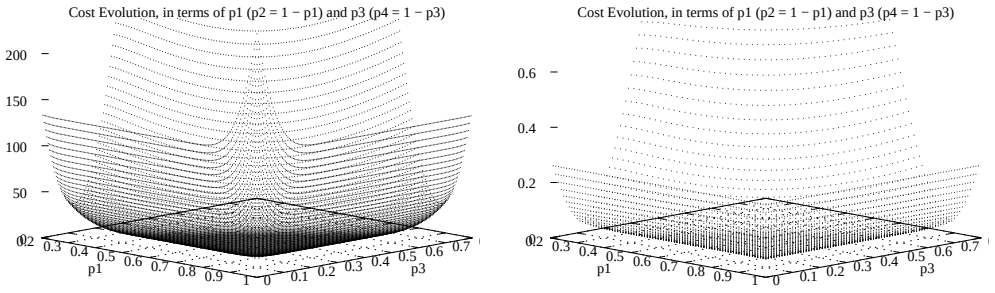
Le critère Min Max

Nous avons également étudié le critère d'optimisation min-max : autrement dit, la répartition de trafic est celle qui minimise la probabilité de perte (taux de pertes dans notre cas) sur le lien où celle-ci est maximale :

$$\min_{\mathbf{p}} \max_{l \in \mathcal{L}} \left(q_{\mathbf{p}}(l) \right) \tag{9.21}$$

La figure 9.8 nous permet de comparer les fonctions de coût du système pour les deux critères cités

précédemment. Dans ce cas $C_{AC} = N10, C_{AB} = N10, C_{BC1} = N10, C_{BC2} = N10$ et $N = 60$. Toutes les sources composantes des agrégats sont caractérisées par leur b.p.e. de Hoeffding avec $\rho = 8$ et $\pi = 20$.



(a) fonction de coût : volume moyen de pertes (b) fonction de coût : min-max

Fig. 9.8: Exemple I : fonctions objectives Volume moyen de pertes et Min Max pour $c_1 = 10, c_2 = 10, c_3 = 10, c_4 = 10$

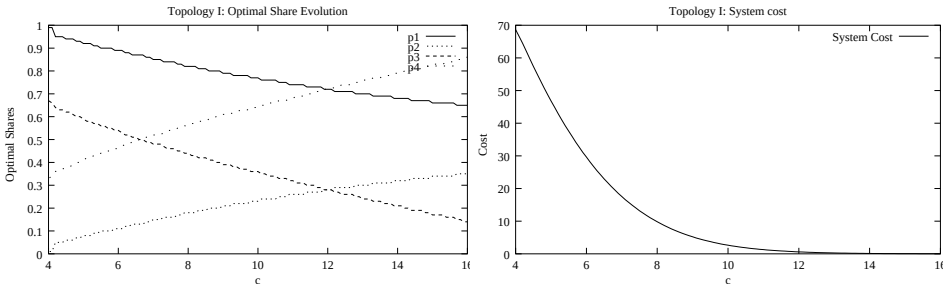
Cette approche présente l'avantage qu'au niveau de l'optimisation numérique (par exemple, par recherche exhaustive), on rejette une solution dès que sur un des liens, le taux de pertes dépasse le candidat actuel. Par contre, au vu des quelques exemples numériques que nous avons réalisées, l'optimum est le même, et la fonction de coût donne moins d'informations sur les variations directionnelles si l'on se trouve dans des répartitions non optimales.

Conclusions de l'exemple : évolution des réseaux

Dans le contexte de l'ingénierie de trafic, on peut s'intéresser à l'impact de l'augmentation de la capacité d'un lien sur le coût du système, ou sur le partage optimal. La figure 9.9 (a) illustre l'évolution de la distribution de trafic optimale (partage) en fonction de la capacité du lien 4 normalisée par N ($c = C_4 N^{-1}$), et (b) l'évolution de la fonction de coût du système. De manière prévisible, l'augmentation de la capacité du lien 4 induit une augmentation du partage p_4, et donc une diminution de p_3. Cette diminution laisse plus de place au LSP 2, qui voit son partage p_2 augmenter. **Ce type d'analyse permettent à l'ingénieur réseau d'identifier les éléments (les capacités des liens) pour lesquels un investissement et une mise à jour ont un impact notable sur la performance du réseau**. Dans l'exemple précédent, l'augmentation de la capacité du lien 4 (C_{BC2}) de $N4$ à $N6$ implique une diminution du coût du système d'environ 50%.

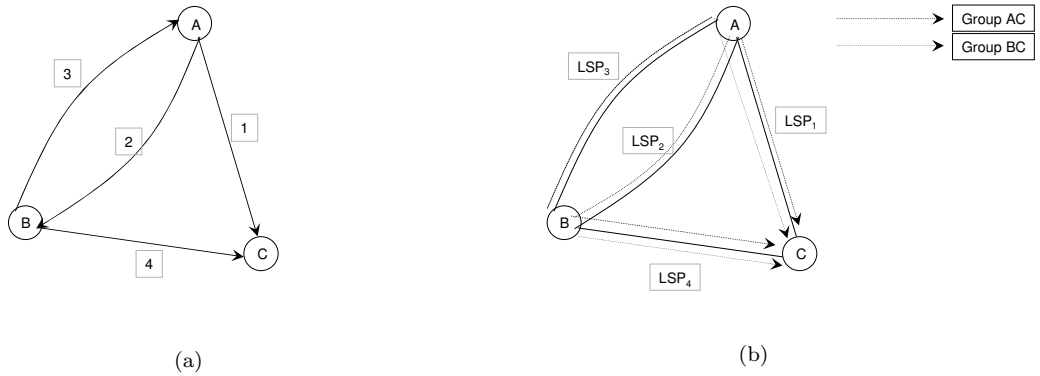
9.5.2. Exemple II

La figure 9.10 illustre la topologie utilisée comme deuxième exemple. Comme dans le cas précédent, un agrégat de trafic doit être acheminé entre A et C, et un deuxième agrégat entre B et C. Les deux agrégats sont constitués comme toujours de N sources de Hoeffding.



(a) (b)

Fig. 9.9: Exemple I : Evolution de la répartition de trafic en faisant varier la capacité du lien 4 (BC2)



(a) (b)

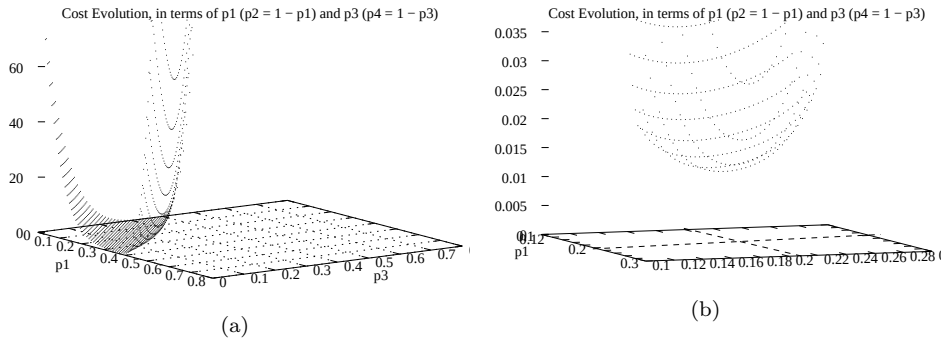
Fig. 9.10: Exemple II : topologie et placement des LSPs. B.P.E utilisant la borne supérieure de Hoeffding.

Les contraintes de stabilité

Dans cette instance du problème nous considérons les capacités normalisées (divisées par N) sont $c_{AC} = 5, c_{AB} = 15, c_{BA} = 15$ et $c_{BC} = 20$. Les sources composant les agrégats ont pour coefficients $\rho = 8, \pi = 21$ et $N = 60$. Le point remarquable dans cet exemple est le fait que la valeur de la capacité du lien 1 (AC) limite fortement le domaine admissible (contrainte de stabilité), comme on peut le voir sur la figure 9.11 : les partages p_1 et p_3, correspondants aux LSPs traversant le lien 1, doivent vérifier :

$$p_1\rho + \rho p_3 q_3 < c_{AC} \tag{9.22}$$

où q_3 est l'équivalent logarithmique du taux de pertes du lien 3 (BA), qui dépend de ρ, π, N, c_{BA} et p_3. Le partage optimal correspond à $(p_1 = 0.2, p_2 = 0.8, p_3 = 0.2, p_4 = 0.8)$ avec un coût de 0.0126066.

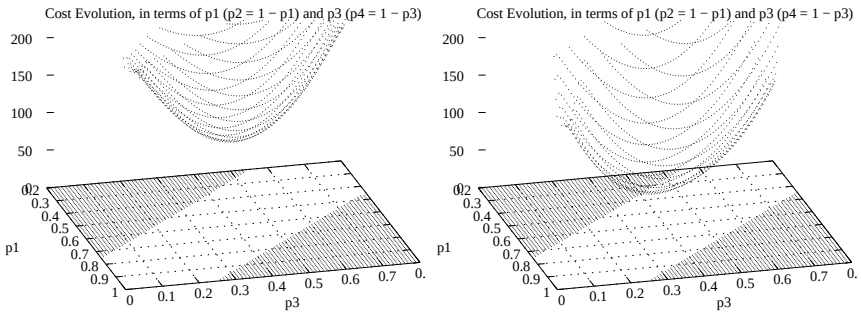


(a) (b)

Fig. 9.11: Exemple II : remarquons l'effet des contraintes de stabilité.

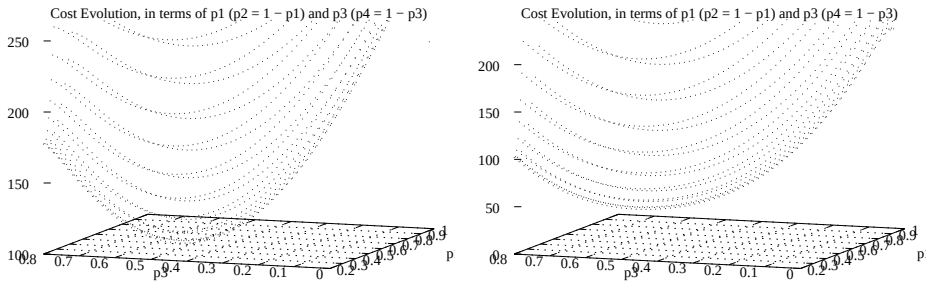
Effet du multiplexage statistique

Considérons un nouvel exemple. Cette fois, toutes les capacités sont $C_i = N10, i = 1..4$, et les sources ont pour coefficients $\pi = 21$ et $\rho = 8$. Remarquons que le système est symétrique. Dans cette configuration, les liens AB et BA deviennent de plus en plus transparents (par rapport aux liens AC et BC, traversés par 2 LSPs). La symétrie du problème nous permet d'attendre, a priori, un optimum pour $p_1 = 0.5$,$p_2 = 0.5$,$p_3 = 0.5$ et $p_4 = 0.5$ (cf. figure 9.12 et figure 9.13). Les zones hachurées correspondent à des valeurs illégales des partages, ne respectant pas les contraintes de stabilité. Comme on peut observer sur la figure (b), plus on augmente N plus le valeur de la fonction de coût à l'optimum est petite, car en déployant de capacités $C_i = N10, i = 1..4$, on néglige dans une certaine mesure l'effet du multiplexage statistique. Autrement dit, grâce à l'effet du multiplexage statistique, pour maintenir un même niveau de performance (un taux de pertes donné), l'augmentation de la capacité nécessaire suite à une augmentation du nombre de sources N est sous linéaire.



(a) $N = 60$ (b) $N = 100$

Fig. 9.12: Exemple II : système symétrique avec $C_i = 10N$ Effet de N

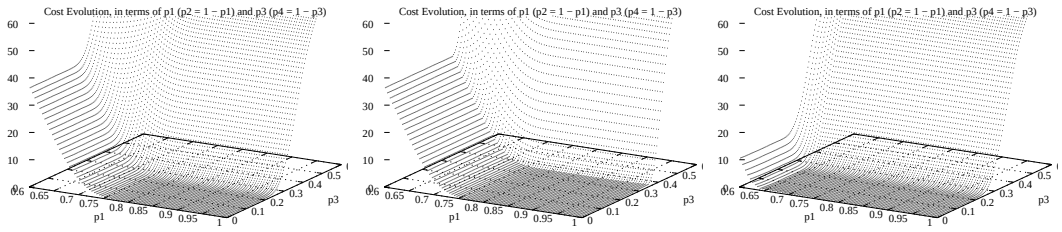


(a) $N = 60$ (b) $N = 100$

Fig. 9.13: Exemple II : effet de N (détail)

Effet des différents Débits Crêtes

Jusqu'ici, nous nous sommes concentrés principalement sur l'effet de paramètres comme les capacités des liens de transmission ou le nombre de sources (N) agrégées. Néanmoins, certaines propriétés stochastiques des sources autres que leur débit moyen jouent un rôle sur les fonctions de coût proposées. Ceci est illustré par la figure 9.14 pour les trois cas suivants : (a) sources à acheminer entre A et C $\pi = 21, \rho = 8$, entre B et C $\pi = 21, \rho = 8$ (b) sources à acheminer entre A et C $\pi = 21, \rho = 8$, entre B et C $\pi = 11, \rho = 8$ et finalement, (c) sources à acheminer entre A et C $\pi = 11, \rho = 8$ entre B et C $\pi = 21, \rho = 8$. Les capacités normalisées (divisées par N) $c_{AC} = 20, c_{AB} = 4, c_{BA} = 4, c_{BC} = 20$, configuration qui intuitivement favorise les LSPs directs (à un seul saut, LSP1 et LSP4). Remarquons comment dans le cas (c), les sources constituant l'agrégat à acheminer entre A et C sont moins sporadiques, et donc la fonction de coût présente un caractère plus plat par rapport à p_1 que les cas (a) et (b).



(a) Sources à acheminer entre A et C $\pi = 21, \rho = 8$, entre B et C $\pi = 21, \rho = 8$

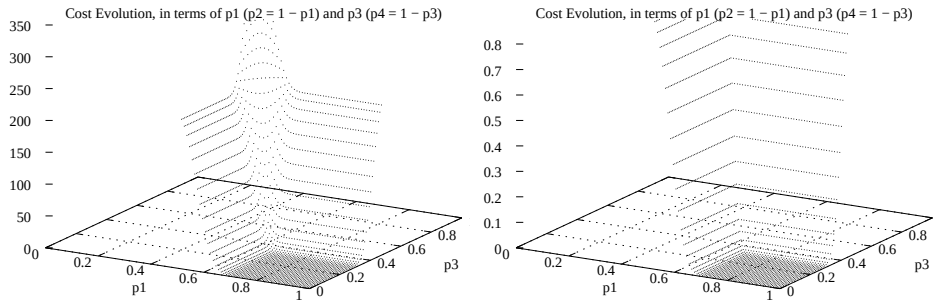
(b) Sources à acheminer entre A et C $\pi = 21, \rho = 8$, entre B et C $\pi = 11, \rho = 8$

(c) Sources à acheminer entre A et C $\pi = 11, \rho = 8$, entre B et C $\pi = 21, \rho = 8$

Fig. 9.14: Exemple II : Différentes fonctions de coût du sytème avec des agrégats à débits moyens égaux, mais avec des débits crêtes différents.

Min-Max

Finalement, nous avons également comparé les deux critères cités, dans la configuration suivante : sources homogènes avec $\pi = 21$ et $\rho = 8$ et capacités normalisées $c_{AC} = 20, c_{AB} = 4, c_{BA} = 4, c_{BC} = 20$ (configuration identique à celle qui nous a permis d'évaluer l'impact des débits crêtes). Le résultat est illustré par la figure 9.15.



(a) Fonction de coût : volume moyen de pertes.

(b) Fonction de coût : min-max.

Fig. 9.15: Exemple II : Différents critères : Minimisation du volume moyen de pertes et min-max

9.5.3. Sources Gaussiennes

Dans le contexte de nos travaux, la caractérisation des sources à l'aide des processus Gaussiens présente l'avantage notable que les expressions dérivées sont relativement simples. Les sources gaussiennes sont caractérisées par leur débit moyen (λ) et par leur variance σ^2. L'expression de leur bande passante effective est la suivante :

$$\alpha(s,1) = \lambda + s\frac{\sigma^2}{2} \tag{9.23}$$

Le point de travail spatial peut être facilement calculé, ainsi que la fonction de taux associée :

$$\begin{aligned} s^* &= \frac{c-\lambda}{\sigma^2} \\ L_r &\asymp e^{-N\frac{1}{2}\frac{(c-\lambda)^2}{\sigma^2}} \end{aligned} \tag{9.24}$$

9.6. Approximations et Simplifications

9.6.1. L'approximation d'invariance

La complexité du problème est réduite sous l'hypothèse que les bandes passantes effectives ne sont pas modifiées par leur passage par des liens de transmission ($\alpha_{in}(s,t) = \alpha_{out}(s,t)$). Bien sûr, la validité de cette approche est fortement mise en cause quand les capacités des liens ont un effet fortement

«tronquant». Cette approximation correspond à une approche pessimiste : en effet, considérer que les fonctions de taux sont invariantes implique que les pertes qui peuvent avoir lieu dans un lien de transmission peuvent être prises en compte par les fonctions de coût plusieurs fois. Une telle approximation est difficilement justifiable, sauf : (a) pour sa simplicité, car l'optimisation devient séparable par lien (les fonctions de coût proposées à chaque lien sont uniquement dépendantes des b.p.e. des points d'entrée) et (b) elle permet de pénaliser *artificiellement* les LSPs de longueur importante.

Simplification par inspection et amincissement du graphe

Tout lien dont la capacité est supérieure à la somme des débits crêtes peut être supprimé.

Séparabilité

Si les groupes de LSPs associés à des couples origine-destination sont à liens disjoints, l'optimisation du réseau est séparable par groupe.

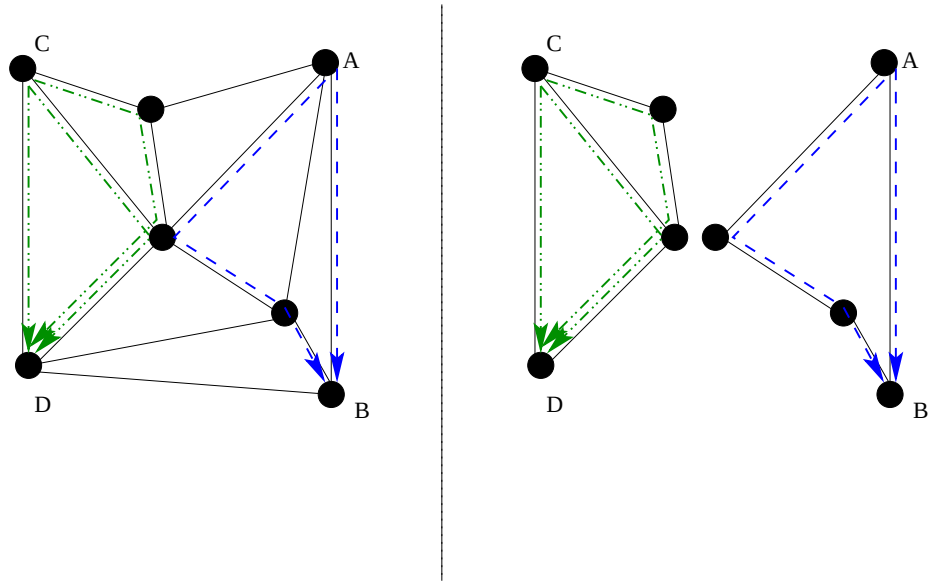


Fig. 9.16: La répartition optimale de trafic est séparable pour les groupes AB et CD. De plus, le groupe AB est "per LSP"-séparable (à liens disjoints)

Approximation des b.p.e.

Définition 9.6.1 (approximation de Padé). L'approximation de Padé d'ordre (m,n) de la fonction f(x) est définie comme la fonction rationnelle p(x)/q(x) avec $\deg(p(x)) \leq m$ et $\deg(q(x)) \leq n$ telle que l'expansion en série de Taylor (ou Laurent) de p(x)/q(x) a un accord maximal avec l'expansion en série de f(x). Typiquement, l'expansion en série est en accord jusqu'au terme au degré m+n.

Il est bien connu que dans le contexte d'un grand degré de multiplexage, avec des débit crêtes largement inférieurs à la capacité nominale du lien, le point de travail spatial tend vers zéro, en considérant des sources de plus en plus caractérisées par leurs débits moyens. Dans de tels cas, il est possible d'approximer les expressions des b.p.e par leur développements en séries de Taylor [53] ou les approximations de Padé. A ce propos, la forme polynomiale ou rationnelle des approximations peut simplifier le traitement numérique des b.p.e. dont l'expression originale est complexe. Pour illustrer ceci, la figure suivante présente la b.p.e. d'un processus de Poisson et son approximation de Padé.

$$\alpha_{\text{Poisson}}(s,t) = \lambda\left(\frac{e^s - 1}{s}\right) \tag{9.25}$$

admet une approximation de Padé [2,1] donnée par :

$$\alpha_{\text{Poisson}}(s,t) \approx \frac{\lambda}{6}\frac{s^2 + 6s + 24}{4 - s} \tag{9.26}$$

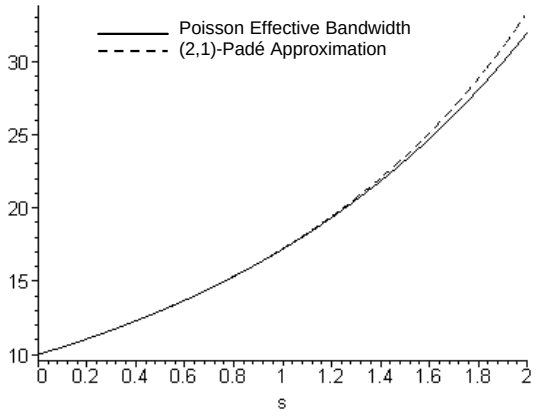


Fig. 9.17: B.P.E d'un processus de Poisson Effective Bandwidth et son approximation (2,1)-Padé Approximation pour un grand degré de multiplexage ($s^* \to 0$)

9.7. Interprétation et Discussion

Rappelons quelques notions de base : l'équivalent logarithmique utilisé, $\asymp e^{-NI}$, dépend du paramètre N, fixé pour une instance donnée du problème et de la fonction de taux, I. Evidemment, la fonction de taux dépendra des propriétés stochastiques des sources (dans les exemples, du débit moyen ρ et du débit crête π, comme illustré par la figure 9.18). Néanmoins, toute fonction de taux vérifie les proprietés de croissance et de convexité par rapport à la différence r entre lepoint où la fonction est évaluée (typiquement la capacité nominale par source du lien, c) et le débit moyen d'un

flot individuel m, et vaut zéro quand $r = 0$[2].

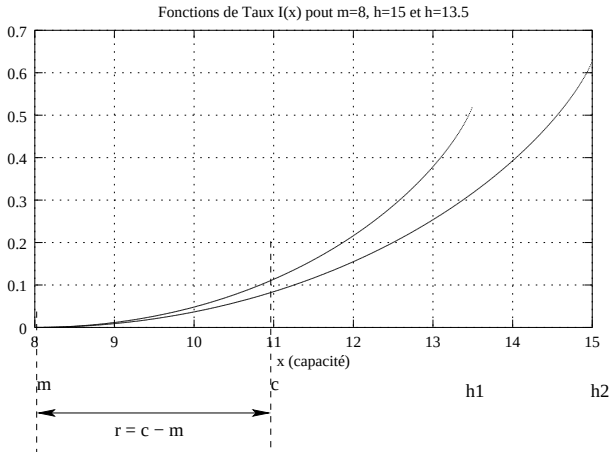


Fig. 9.18: Exemples de fonctions de taux. Remarquons les propriétés de croissance, de continuité et de convexité.

Autrement dit, la fonction de taux correspond à *une fonction d'utilité* avec de très bonnes propriétés (continuité, convexité, croissance) en fonction de *la bande passante résiduelle moyenne par source.* Cette affirmation nous permet de considérer l'approche suivante : la répartition de trafic doit «éloigner» la bande passante moyenne offerte par source de la capacité par source du lien, c.à.d. *maximiser la capacité résiduelle moyenne par source.* Dans le contexte d'un réseau, examinons le critère d'optimisation suivant :

$$\max_{\mathbf{p}} \min_{l \in \mathcal{L}} \left(c(l) - \lambda(l) \right) \tag{9.27}$$

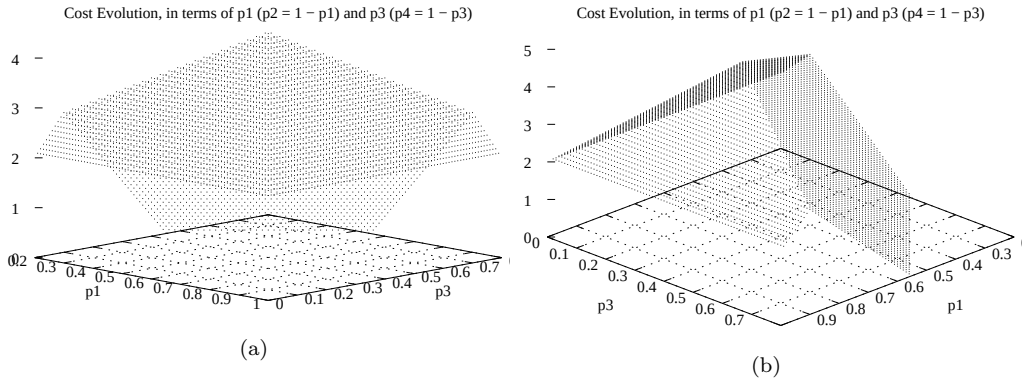
où

$$\begin{aligned} \lambda(l) &= \lim_{s \to 0} \sum_{\substack{m \in \mathcal{M} \\ l \sim m}} \prod_{j=1}^{i_m(l)-1} q(l_m^j) \alpha_m \left(\prod_{j=1}^{i_m(l)-1} q(l_m^j) s, t \right) \\ &= \sum_{\substack{m \in \mathcal{M} \\ l \sim m}} \prod_{j=1}^{i_m(l)-1} q(l_m^j) p_m \left(\lim_{s \to 0} \gamma_{g(m)}(s,t) \right) \end{aligned} \tag{9.28}$$

car pour un lien l, $\lambda_{out} = q(l)\lambda_{in}$. Nous avons toujours besoin de l'asymptotique du grand nombre d'usagers pour le calcul du taux de pertes sur chaque lien. La figure 9.19 montre la fonction d'utilité *(ang. benefit function)* pour la topologie utilisée dans l'exemple I, avec $C_i = N10, i = 1..4$, $N = 60$ et $\pi = 20, \rho = 8$. Remarquons que la répartition optimale de trafic reste la même (cf. figure 9.8).

Si nous faisons a priori l'hypothèse que les taux de pertes de chaque lien sont négligeables (c.à.d. que $q(l) \approx 1$) en gardant les contraintes de stabilité, ceci nous permet de formuler le problème

[2] Ceci est vrai car la fonction de taux est la transformée de Fenchel-Legendre (transformée convexe) de la transformée de LogLaplace du processus des accroissements du trafic $L_{x[0,1)}(\theta)$. La transformée convexe atteint son minimum en $x = \frac{L_{x[0,1)}(\theta)}{\partial\theta}|_{\theta=0}$, c.à.d $\mathbb{E}\left(x[0,1)\right)$. La valeur en ce point est zéro.



(a) (b)

Fig. 9.19: Maximisation de la capacité résiduelle par source, topologie de l'exemple I.

d'optimisation suivant :

$$\max_{\mathbf{p}} \min_{l \in \mathcal{L}} \left(c(l) - \sum_{\substack{m \in \mathcal{M} \\ l \sim m}} p_m \lambda_{g(m)} \right) \tag{9.29}$$

Problème plus simple, nous permettant de faire le lien avec des travaux similaires de la théorie de graphes. Bien évidemment, une fois l'optimum calculé, une estimation des taux de pertes est nécessaire afin de valider l'hypothèse $q(l) \approx 1$.

9.8. Conclusions

De façon générale, les conclusions que nous tirons des analyses ici réalisées sont les suivantes :

Effet du multiplexage statistique : Les principes de grandes déviations qui nous permettent de quantifier la décroissance exponentielle de la probabilité de pertes ou de la queue de distribution du travail cumulé dans une file d'attente reflètent les bénéfices du multiplexage statistique. L'analyse effectuée nous permet de quantifier le gain que ceci représente (par opposition aux critères d'allocation statique).

Unicité de la solution : Même si dans certains cas on peut prouver (d'un point de vue mathématique) l'unicité de l'optimum, ceci n'est certainement pas vrai en général : l'exemple le plus simple correspond à l'ensemble des répartitions de trafic pour lesquelles les débit crêtes des flots offerts à un lien sont inférieurs à la capacité par source du lien. Cet ensemble des répartitions est optimal avec un coût nul.

Evaluation des Etats Sous Optimaux : Il se peut que la technologie utilisée limite la granularité du partage de charge (par exemple, des mécanismes de hachage avec une résolution de 0.1). L'approche que nous proposons ici nous permet de calculer l'optimum théorique mais aussi la dégradation de performance d'un état sous optimal par rapport à l'optimum.

Ingénierie de Trafic et Dimensionnement : A l'aide des études comme celle illustrée sur la figure 9.9 les opérateurs de réseaux peuvent évaluer le gain de performance qu'apporte un investissement (ici, une augmentation de la capacité d'un lien) et prendre des décisions en tenant compte de contraintes économiques et technologiques.

Validité des approches simplistes : Le nombre de topologies et de cas particuliers analysés reste néanmoins limité. Nous n'avons pas évalué sa complexité dans des réseaux de taille. Nous pouvons cependant affirmer que sa complexité dépendra fortement de la disposition ou layout des LSP (un «degré» ou nombre de LSPs traversant un lien quelconque augmente les inter-rélations entre les différents taux de pertes) et de la vitesse de calcul des fonctions de taux, qui peuvent devenir difficilement calculables pour des réseaux de grande taille. De ce fait, les partages optimaux peuvent être approximés à l'aide des approches simplistes. En résumé : grâce à l'effet du multiplexage statistique et si le dimensionnement du réseau respecte les contraintes de stabilité, il peut suffire d'éviter les liens fortement chargés. Le caractère plat des fonctions de coût vient confirmer cette approche. Les débordements ont lieu rarement sauf si l'on se trouve dans des états «presque critiques». Dans ce cas-là, l'approche proposée ici devient intéressante vis-à-vis des performances du réseau.

L'approche proposée intègre les propriétés statistiques du trafic, et pénalise fortement les liens dont la charge est proche de 1.

Extensions et Perspectives

La fonction objective présentée dans ce chapitre ne prend pas en compte des facteurs comme la longueur des chemins ou les délais de propagation. Nous pensons qu'il est possible de généraliser les fonctions de coût en ajoutant des termes qui pénalisent la longueur relative d'un chemin par rapport à celle des autres chemins appartenant au même groupe, mais cette question reste ouverte.

La *transformation* de la bande passante effective d'un flot par son passage par une file d'attente avec un buffer de taille non négligeable (de l'ordre de la capacité de transmission) reste une question ouverte qui rend difficile le développement d'un modèle plus complet. Cette transformation dépendra non seulement des propriétés du flot (de son histoire), mais également de son contexte de multiplexage (les autres flots externes, ainsi que la taille du buffer et la capacité). Nous sommes restés dans le contexte sans buffer, et nous avons caractérisé la b.p.e. de sortie en utilisant une approche pragmatique.

Remerciements

Nous remercions le Prof. Ravi Mazumdar et le Dr. Anthony Busson pour leurs commentaires et leurs remarques, ainsi que le Prof. Ness Shroff dont un article à apparaître [30] nous a permis de mieux comprendre les enjeux de l'extension aux réseaux de l'asymptotique du grand nombre d'usagers.

Troisième partie .

Conclusions Générales et Perspectives

10. Conclusions et Perspectives

Le dimensionnement, l'évaluation de performances et l'optimisation d'un réseau nécessitent une caractérisation du trafic pouvant modéliser les propriétés statistiques des sources. Cette caractérisation est complexe du fait, par exemple, des phénomènes observés de dépendance longue et d'auto similarité. Ces propriétés rendent certaines approches classiques comme l'utilisation de processus Markoviens non réalistes, et donc, inefficaces.

Dans cette thèse, nous souhaitons proposer une approche de dimensionnement du partage de charge générique qui puisse s'appliquer pour tout type de trafic : nous avons donc utilisé une modélisation en files d'attente pour laquelle le trafic est représenté de manière générique par sa bande passante effective. La fonction de coût que nous optimisons prend en compte des paramètres qui reflètent les besoins des réseaux opérationnels. A titre d'exemple, nous considérons des quantiles des mesures de performance et non pas des moyennes. Le modèle ainsi défini est très complexe ; il n'en existe pas de solutions exactes. Du fait que nous nous intéressons au cœur du réseau, il est réaliste de regarder des solutions asymptotiques dans le nombre de flux transportés. Nous avons donc fait appel à la théorie des grandes déviations.

A ce propos, nous utilisons la théorie des bandes passantes effectives qui s'avère générique. Le nombre de travaux portant sur cette théorie est relativement important, mais la majorité de ceux-ci sont centrés sur l'asymptotique de grand buffer, qui présente une structure moins riche que celle de l'asymptotique de grand nombre d'usagers. Cette asymptotique est valide sous l'hypothèses d'un grand nombre de sources stationnaires, hypothèse largement validée dans le coeur d'un réseau opéré, où l'effet du multiplexage statistique est facilement observable. De plus, on observe une stationnarité du trafic étant au moins locale pendant une certaine fenêtre de temps.

Grâce à la puissance de cette théorie nous avons pu obtenir l'expression analytique des fonctions de coût étudiées en fonction des règles de partage, que par la suite nous avons optimisées.

10.1. Synthèse de notre contribution

Cette thèse avait pour objectif d'approfondir notre compréhension des problèmes de partage de charge et de répartition de trafic en général. Nous avons proposé des modèles qui représentent fidèlement les réseaux réels. Nous en avons évalué les performances, nous avons identifié les paramètres les plus significatifs de ces modèles (c'est-à-dire ceux ayant un impact important sur ces performances) et nous avons obtenu à partir de là des règles de dimensionnement simples.

Dans des cas particuliers du point de vue mathématique mais généraux du point de vu de l'ingénierie du trafic, nous avons pu déduire des règles de partage génériques. Par exemple, dans le cas ou on désire borner le délai dans tous les chemins retenus pour le partage par la même borne, nous avons montré que le partage optimal est celui qui attribue à un chemin un pourcentage du flux total proportionnel à la capacité du chemin (la constante de proportionnalité étant 1 sur la somme des

capacités des chemins retenus pour le partage).

Quand nous parlons de chemin, il peut s'agir de liens physiques entre deux équipements ou de LSPs entre deux équipements. Dans le deuxième cas, la capacité des chemins peut varier dans le temps. En effet, il se peut qu'il n'y ait une allocation de ressources préalable sur ce type de chemins. Nous avons donc étendu le modèle pour prendre en compte ce phénomène. Nous avons défini un concept dual de la bande passante effective, que nous avons appelé la capacité effective, pour représenter de manière générique la variation de la capacité.

Sur la base des résultats obtenus, nous avons mis en évidence qu'un dimensionnement basé sur une capacité moyenne peut s'avérer significativement sous optimal et nous avons quantifié cela.

Afin de proposer une approche d'ingénierie de trafic adaptée au cas de figure étudié, nous avons analysé l'impact de diverses méthodes d'évaluation «on-line» de la capacité effective inspirées des méthodes connus d'estimation de la bande passante effective. Nous avons conclu que les méthodes les plus simples, basées sur l'estimation du débit minimum, débit crête et débit moyen s'avèrent largement suffisantes. Ces idées rendent viable une approche d'ingénierie de trafic adaptative, dans laquelle on fait évoluer le partage en fonction des mesures réalisées sur le réseau.

Dans la dernière partie de ce travail nous avons regardé le réseau dans sa globalité, c'est-à-dire que nous avons modélisé les interactions entre les partages de charge entre l'ensemble des couples entrée-sortie. Cela nous permet de mieux optimiser le point de fonctionnement du réseau mais au coût d'une grande complexité des calculs nécessaires.

En synthèse, nous nous sommes intéressés à l'optimisation des politiques partage de charge dans un réseau supportant le routage à la source, tel qu'un réseau MPLS. Une modélisation générique en files d'attente alimentées par un trafic caractérisé par sa bande passante effective et l'utilisation de la théorie des Grandes Déviations nous a permis de déduire de règles d'ingénierie génériques pour le partage de charge dans divers contextes et notamment quand la capacité des liens varie. Une approche d'ingénierie de trafic adaptative a pu ainsi être proposée. Des propriétés structurelles sur les politiques optimales ont été démontées pour des cas particuliers mais importants dans le domaine de l'ingénierie du trafic.

Le déroulement détaillé de cette thèse a été le suivant :

1. Proposition de modèles génériques pour le partage de charge, avec une caractérisation simple des b.p.e. "offertes à chaque route" en fonction des b.p.e. des agrégats d'entrée. Obtention des conditions d'optimalité locale et globale du partage et des règles de répartition. Illustration de l'approche à l'aide de modèles de trafic simples. (Chapitre 6).
2. Extension du modèle pour prendre en compte la variabilité temporelle de la capacité d'une connexion de bout en bout. Définition de la notion de "capacité virtuelle effective". Illustration et quantification des limitations et de la dégradation de performances des approches classiques dans un contexte à capacité variable. (Chapitre 7, première partie).
3. Proposition et évaluation des performances d'un mécanisme de partage de charge adaptatif, avec calcul dynamique de la répartition optimale du trafic. Ce mécanisme estime les b.p.e. et les capacités effectives à partir de traces et prend en compte la non-stationnarité globale observée dans les systèmes réels. Comparaison avec la méthode de Dembo et validation de l'approche par simulation. (Chapitre 7, deuxième partie).

4. Extension du modèle à une optimisation globale sur l'ensemble du réseau. L'étude se base sur des résultats récents de la théorie des grandes déviations. Proposition d'une heuristique pour la répartition du trafic. Illustratrion à l'aide des bornes de Hoeffding des b.p.e. (chapitres 8 et 9).

10.2. Conclusions

Nous présentons dans ce qui suit les conclusions des différents études que nous avons réalisées.

10.2.1. Partage de charge

Les conclusions que nous tirons de l'analyse du problème de partage de charge peuvent être résumées dans les affirmations suivantes :

1. La prise en compte des propriétés stochastiques des sources constituant l'agrégat de trafic est nécessaire pour en quantifier l'effet sur ces paramètres intéressants tels que la probabilité de débordement ou le taux de pertes. Nous avons montré et quantifié que des modèles de trafic différents peuvent avoir des impacts différents sur la dégradation de performance du système dans des états sous - optimaux : par exemple, sous les contraintes de stabilité, un partage non optimal avec des sources à caractère très sporadique représente une dégradation des performances du réseau beaucoup plus notable que celle causée par des sources à caractère plus constant.
2. Bien evidemment, des critères de performance différents impliquent de partages optimaux différents. L'exemple classique est le suivant : si le critère consiste à minimiser le délai moyen, à faible charge le trafic sera acheminé dans sa totalité sur le chemin ayant la capacité la plus élevée. Par contre, pour une famille importante de problèmes d'optimisation, même à faible charge le trafic est réparti de façon proportionnelle aux capacités nominales.
3. En ce qui concerne le partage de charge optimal, certaines configurations (par exemple, le dimensionnement des buffers et des capacités assurant un délai maximal constant) présentent une insensibilité aux propriétés stochastiques des sources. Dans de tels cas, le partage optimal correspond à des critères intuitifs et pragmatiques comme la proportionnalité à la capacité nominale.
4. Le modèle proposé peut être étendu de façon immédiate dans le contexte d'un réseau à classes de service si l'on considère que le système attribue une capacité constante à l'agrégat de chaque clase. Cette hypothèse correspond à une approche «pire cas» par rapport à un système réel dans lequel les ordonnanceurs redistribuent la capacité non utilisée par une classe de trafic aux autres classes.

10.2.2. Partage de charge à capacité variable

1. Les résultats du chapitre 7 sur le partage de charge correspondent à un modèle plus riche intégrant la variabilité temporelle qui pourrait caractériser la capacité associée à un canal de bout en bout établi sans réservation stricte de ressources.

2. Les moments de deuxième ordre et d'ordres supérieur des processus stochastiques modélisant les différents processus de service ont un impact fortement négatif sur la performance du réseau. Ceci explique l'insuffisance des approches basées sur des moyennes : il est toujours préférable d'acheminer du trafic sur des liens à caractère stable que sur des liens pour lesquels à moyenne égale la capacité instantanée présente de fortes oscillations temporelles.
3. L'utilisation des bornes de Hoeffding pour l'estimation des capacités et des bandes passantes effectives est suffisante pour proposer un mécanisme de partage de charge adaptatif, ayant comme objectif l'adaptation automatique de la répartition optimale du trafic aux changements des propriétés stochastiques du système. En effet, les propriétés stochastiques d'un système réel ne sont stationnaires que sur des intervalles de temps de durée limitée.

10.2.3. Extensions aux réseaux

1. La caractérisation *simple* des processus de sortie, au moins sous des régimes asymptotiques, et l'identification des conditions sous lesquelles les résultats asymptotiques ayant une portée locale peuvent être étendus, au moins jusqu'à un certain degré, à un contexte de réseaux de files d'attente restent aujourd'hui les questions ouvertes les plus contraignantes pour la proposition d'un modèle général réaliste.
2. L'extension d'une approche relativement *microscopique* à un réseau de taille importante présente certainement des problèmes de complexité limitant son applicabilité.
3. Le problème de la répartition de trafic dans un réseau de grande taille, problèmes important dans le cadre de l'ingénierie de trafic, admet plusieurs solutions satisfaisantes. En termes très simplifiés (et qui nous pensons n'étonneront personne), les modèles étudiés nous permettent d'énumérer quelques règles simples de dimensionnement :

   Stabilité La répartition de trafic doit respecter les contraintes de stabilité.

   Tout se passe bien à faible charge Il faut éviter les liens à très forte charge, avec des débits moyens très proches des capacités nominales des liens.

   Passage à l'échelle Des approches classiques de la théorie de graphes visant à la maximisation des bandes passantes moyennes sont complémentaires à nos travaux et sont intéressantes lorsque la taille du réseau peut limiter l'applicabilité des approches microscopiques.

10.3. Extensions Proposées

Caractérisation du trafic offert : En ce qui concerne le partage de charge et les études présentées dans cette thèse, nous avons choisi la solution qui consiste à caractériser le trafic offert à un ressource (LSP, file d'attente, etc.) en utilisant une approche fluide, à l'aide de la notion de «partage» $p \in [0, 1]$. D'autres modèles peuvent être proposés à l'aide des familles de fonctions de «routage». Le principal inconvénient vient du fait qu'il est difficile de caractériser la bande passante effective offerte à partir de la connaissance de la bande passante effective de l'agrégat et d'un nombre réduit de paramètres.

Evaluation des Méthodes d'Optimisation : Les problèmes d'optimisation Lagrangienne présentés dans cette thèse sont rarement résolubles analytiquement et nécessitent l'utilisation de méthodes numériques pour le calcul des valeurs optimales. Des études concernant la convergence et d'autres propriétés de certaines méthodes d'optimisation numérique telles que le Recuit Simulé ou la méthode Tabou, peuvent apporter des améliorations notables aux travaux présentés ici.

Partage de charge : La problématique du partage de charge soulève encore de nombreuses questions. Les modèles proposés peuvent intégrer le partage de charge hiérarchique, mais ils supposent souvent une connaissance parfaite du système. Il devient nécessaire de caractériser leur faisabilité dans des systèmes à connaissance imparfaite, ou avec des délais de propagation pouvant entraîner le calcul des différents paramètres en utilisant des informations imprécises ou obsolètes, et pouvant être la cause de phénomènes d'oscillation.

Partage de charge orienté connexion : Nous nous sommes focalisés sur la caractérisation du partage de charge «au niveau du paquet». Néanmoins, il est également intéressant d'étudier la caractérisation du partage de charge et son évaluation de performances au niveau des flots. Dans cette optique, la théorie des flots stream et des flots élastiques pourrait constituer un bon point de départ. De plus, nous avons souvent utilisé des hypothèses d'indépendance dans nos modèles, donc ne prenant pas en compte les dépendances et les interrelations entre les points terminaux comme ceux participant à une session TCP commune.

Solutions adaptatives ; intégration avec le re-routage et la protection par commutation : Une première approche pour le partage de charge adaptatif a été présenté au chapitre 7, approche permettant, dans le cas extrême, l'adaptation du système suite à l'occurrence d'une panne. Nous avons déjà mentionné comment le partage de charge et la protection par commutation sont des concepts très liés (la protection par commutation correspond à une adaptation du partage pour laquelle la proportion de trafic associé à un chemin non fonctionnel est nulle). Néanmoins, de nombreuses questions restent encore ouvertes concernant l'intégration des mécanismes de partage de charge et des mécanismes de protection et de restauration basés sur MPLS. Rappelons que ces derniers nécessitent d'assurer des temps de restauration de l'ordre de la ms, et que la synchronisation entre les différentes couches protocolaires n'est pas un problème simple. Cependant, nous pensons que ces deux aspects sont fortement liés et que de nouvelles études présentant une approche unifiée seraient intéressantes voire nécessaires.

Solutions adaptatives ; intégration avec la Métrologie : L'intégration entre des travaux concernant la métrologie et nos travaux a été souvent évoquée dans le présent document (citons, par exemple, l'estimation des bandes passantes effectives à partir des traces, ou comment l'application de techniques de type «packet-pair» peuvent donner des estimations sur les capacités associées à des canaux IP / MPLS). Nous sommes convaincus que la métrologie est un axe de recherche important complémentaire aux résultats théoriques développés dans cette thèse.

Extensions à des architectures à différentiation de services : Les modèles proposés dans les chapitres 6 et 7 sont susceptibles d'être étendus pour prende en compte des classes de trafic dans une architecture à différentiation de services. Cette extension est justifiée par l'intérêt croissant des opérateurs réseaux pour déployer des solutions intégrant les architectures MPLS et Diffserv. Certains résultats existants permettant d'approximer la probabilité de certains événements dans des files à priorités [53] peuvent s'avérer bien adaptés. Par ailleurs, l'extension de nos travaux à différentes

classes de trafic est relativement simple si l'on fait l'hypothèse qu'une capacité fixe est allouée à chaque classe. Ceci correspond à une approche "pire cas" dans le contexte d'une gestion de files d'attente où, selon la politique de service mise en place, la capacité non utilisée par une classe est utilisée par les autres classes.

Quatrième partie .

Annexes

A. Techniques des Grandes Déviations

A.1. Motivation

Cette annexe présente les éléments essentiels de la théorie des grandes déviations[1], appliquée à la théorie des files d'attente. La théorie fournit les bases de celle des bandes passantes effectives présentée au chapitre 5. L'annexe est structurée de la façon suivante : On justifie la notion de grande déviation par rapport à la moyenne empirique d'une suite de variables aléatoires, en montrant l'insuffisance de la loi des grandes nombres et du théorème central limite. Ensuite, la borne de Chernov, cas particulier de l'inégalité de Bienaymé - Chebychev, donne les bases intuitives des résultats de la théorie des grandes déviations. Le théorème de Cramer et son extension de Gartner-Ellis sont présentés avant de définir la notion de Principe de Grandes Déviations ou PGD (le terme anglais *Large Deviation Principle ou LDP* est aussi souvent utilisé), terme clé dans la théorie. Le principe de contraction est présenté ensuite, permettant l'obtention des PGDs pour des suites de v.a à partir d'un PGD établi pour une suite, sous certaines conditions techniques. Finalement, on présente les résultats les plus importants pour l'estimation de la probabilité de certains événements dans le contexte des files d'attente, notamment *l'asymptotique du grand buffer* et *l'asymptotique du grand nombre d'usagers.*

Tout travail concernant la théorie des grandes déviations se doit de citer l'ouvrage "Large Deviations Techniques and Applications" [23]. Cet ouvrage est un exposé rigoureux de la théorie des grandes déviations. L'objectif de cette annexe est d'en présenter de façon simple les résultats les plus utiles à des lecteurs n'ayant pas une formation probabiliste approfondie, afin de faciliter la compréhension du contenu de cette thèse. La présente annexe ne se veut pas une référence exhaustive ni rigoureuse de la théorie des grandes déviations. Le lecteur intéressé est invité à consulter les références citées pour un exposé détaillé de cette théorie et pour les preuves des résultats présentés.

A.2. Introduction

La théorie des grandes déviations a trait au calcul et à l'estimation de la probabilité de certains événements considérés comme rares. La communauté scientifique du télétrafic a effectué ses premiers travaux concernant la théorie des grandes déviations pour étudier certains événements d'intérêt pour l'ingénierie de trafic, comme le débordement d'un seuil (sauf si la taille du buffer a été dimensionnée avec des contraintes de délai) qui correspondent bien à l'idée d'événement "rare". Hui [47] a été un des premiers auteurs à établir un lien entre le télétrafic et les grandes déviations. Il est parfois utile de définir la théorie des grandes déviations comme l'étude de certains événements *sous un régime asymptotique déterminé* (autrement dit quand un paramètre du système est suffisamment grand).

[1] Le contenu de cette annexe est basé sur les notes de cours DEA Probabilités Appliqués, module Processus Ponctuels et Applications, dont les responsables sont M. François Bacelli et M. Laurent Massoulié, ainsi que sur différents ouvrages, notamment [23] et [13]

Nous reviendrons sur cette idée, notamment dans les résultats connus comme l'asymptotique du grand nombre d'usagers ou l'asymptotique de grand buffer.

La théorie des grandes déviations est assez récente. Comme nous allons voir dans la suite, c'est vers 1938 que le théorème de Cramer [21] a établi les bases pour l'étude d'événements "s'éloignant de la moyenne", mais ce n'est qu'en 1960 que Bahadur et Rao ont publié leurs résultats, et qu'à partir des années 1980-1990 que le nombre de travaux sur les grandes déviations est devenu important.

Quel est l'objectif de la théorie des grandes déviations ? L'exemple souvent cité dans les introductions à cette théorie est le suivant : si l'on réalise 300 tirages de pile ou face, il est justifié de considérer que l'on obtiendra environ 150 piles. On peut également s'intéresser à la probabilité d'obtenir plus de 270 faces... Un deuxième exemple qui présente plus d'intérêt pour l'ingénieur de trafic est le suivant : le multiplexage statistique typique des réseaux à commutation de paquets nous permet de dimensionner des systèmes en tirant parti de la sporadicité des sources. En effet, la bande passante non utilisée par une source peut être utilisée par une autre. On n'est ainsi pas obligé de dimensionner le système avec une approche "pire cas" qui gaspillerait des ressources. Néanmoins, en accord avec la nécessité de donner des bornes stochastiques et de définir des contrats de QoS à l'aide de paramètres mesurables, il est intéressant de pouvoir évaluer la probabilité (certes rare) que toutes les sources *émettent en même temps,* et que le système déborde. De plus, on s'intéressé également à la trajectoire des processus la plus probable donnant lieu à un tel événement *(ang. most likely path).*

A.2.1. Une introduction par l'exemple : Multiplexage Sans Buffer

Considérons un système à temps discret. A chaque instant de temps, un certain nombre $N \in \mathbb{N}$ de sources génèrent un trafic avec une certaine distribution. On fait l'hypothèse que les sources suivent une distribution commune et sont mutuellement indépendantes (sources i.i.d.). Ainsi, le trafic généré par une source i à l'instant t est modélisé comme une variable aléatoire que l'on note X_i. L'évaluateur de performances décide, dans une première étape et connaissant a priori le nombre de sources N, de mettre en place un lien de transmission de capacité totale $C = Nc$ (on parle souvent de c comme la *capacité par source*). Etant donné l'aspect aléatoire des sources, la seule façon d'assurer qu'à tout instant la totalité du trafic généré sera servi consiste à dimensionner le système de sorte que la capacité par source soit plus grande que le débit crête instantané d'une source. Mathématiquement, on parle du *Supremum Essentiel ou* ess sup, correspondant à la généralisation de l'idée de maximum pour des fonctions mesurables, à la différence près que les valeurs d'une fonction dans un ensemble de mesure nulle n'affectent pas le ess sup.

Définition A.2.1 (Supremum Essentiel). Soit E un espace mesuré, de mesure μ et soit f une fonction mesurable $f : E \to \mathbb{R}$ Le ess sup est la plus petite valeur a telle que l'ensemble $\{x : f(x) > a\}$ soit de mesure nulle. Si cette valeur n'existe pas le essup est ∞. Autrement dit, pour une v.a. X on note $\bar{X} \triangleq \text{esssup}\,(X) = \sup\{x : \mathbb{P}\,(X > x) > 0\}$

Même si dans la pratique toute source physique de trafic a un essup fini, cette approche peut s'avérer extrêmement coûteuse. La réponse à ce problème consiste à prendre en compte le phénomène de multiplexage statistique. On souhaite, bien sûr, dimensionner le système de telle sorte que $\mathbb{E}[X] < c$ afin d'avoir un système où les débordements sont "rares" ou "non souhaitables" tout en respectant la stabilité. On peut faire ici quelques remarques :

– **a) Evaluation***Pour une capacité totale fixe $C = Nc$, si l'on note $S_N = \sum_{i=1}^{N} X_i$ le trafic total offert par l'agrégat des sources, quelle est la performance du système ?*. Plus précisement, *Quelle est la probabilité qu'à un instant de temps t il y ait des pertes ?* ou *Quelle est le taux de pertes par unité de trafic ?*. Ces deux mesures de performance peuvent s'exprimer de la façon suivante :

$$\mathbb{P}\left(\sum_{i=1}^{N} X_i[n] \geq Nc\right) = \mathbb{P}\left(S_N \geq Nc\right) \tag{A.1}$$

$$\mathbb{P}_L = \frac{\mathbb{E}[(S_N - Nc)^+]}{\mathbb{E}[S_N]} \tag{A.2}$$

– **b) Dimensionnement** *Si l'on fixe un critère de performance (par exemple une probabilité de perte) γ comme objectif, quelle est la valeur de C minimale à mettre en place ? ou, si C est fixe, quel est le nombre maximal de sources que l'on peut multiplexer ?*

Répondre à ce type de questions est un des objectifs de la théorie des grandes déviations appliquées à l'évaluation des performances des réseaux.

Dans une première étape, on cherche à caractériser la probabilité donnée en (A.1) et illustrée par la figure A.1. Si l'on note $M_N = \frac{\sum_{i=1}^{N} X_i}{N}$ (moyenne empirique), la probabilité (A.1) devient la caractérisation de la queue de distribution de la v.a. M_N

$$\mathbb{P}\left(M_N \geq c\right) \tag{A.3}$$

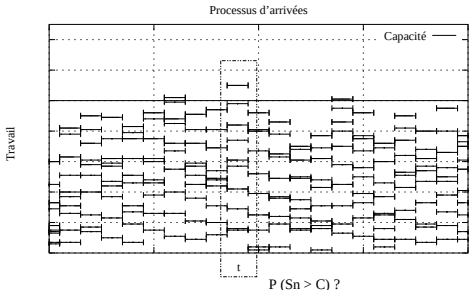


Fig. A.1: Introduction : Evaluation de la Probabilité de Perte

Rapport avec le théorème central limite On peut faire appel à des outils mathématiques pour répondre à ce type de questions : la Loi des Grandes Nombres, qui établi que la moyenne empirique d'une suite de variables aléatoires i.i.d. tend presque sûrement vers sa moyenne, l'inégalité de Bienaymé-Chebytchev, et le théorème centrale limite. Soit X une v.a. de variance finie. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables v.a. indépendantes de même loi que X. De façon informelle, la loi faible des grands nombres dit que

$$\begin{gathered} \frac{1}{N}\sum_{i=1}^{N} X_i \to E[x] \quad p.s. \\ \mathbb{P}\left[\left|\frac{X_1 + \cdots + X_N}{N} - \mathbb{E}[X]\right| > \varepsilon\right] \to 0 \quad \text{quand } N \to \infty \quad \forall \varepsilon > 0 \end{gathered} \tag{A.4}$$

Bien sur, les relations suivantes sont aussi vérifiées :

$$\begin{aligned} \mathbb{E}\left[\frac{X_1+\cdots+X_N}{N}\right] &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[X_i] = \mathbb{E}[X] \\ Var\left[\frac{X_1+\cdots+X_N}{N}\right] &= \frac{1}{N^2}\sum_{i=1}^{N}Var[X_i] = \frac{1}{N}Var[X] \end{aligned} \tag{A.5}$$

l'inégalité de Bienaymé-Chebytchev dit que

$$\mathbb{P}\left[\left|\frac{X_1+\cdots+X_N}{N} - \mathbb{E}[X]\right| > \varepsilon\right] \leq \frac{Var[X]}{N\varepsilon^2} \tag{A.6}$$

Une conséquence classique de ces résultats est que l'erreur commise en approximant $\mathbb{E}[X]$ par sa moyenne empirique $N^{-1}(X_1+\cdots+X_N)$ est de l'ordre de $1/\sqrt{N}$. Si $\varepsilon = \frac{c\sqrt{Var[X]}}{\sqrt{N}}$

$$\mathbb{P}\left[\left|\frac{X_1+\cdots+X_N}{N} - \mathbb{E}[X]\right| > \frac{c\sqrt{Var[X]}}{\sqrt{N}}\right] < \frac{1}{c^2} \tag{A.7}$$

Rapellons (T.C.L.)

$$\lim_{N\to\infty}\mathbb{P}\left(\frac{\sum_{i=1}^{N}X_i - N\mathbb{E}[X]}{\sqrt{n\,\mathrm{Var}(X)}} \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt \tag{A.8}$$

L'évaluateur de performances peut essayer d'appliquer le T.C.L. lorsque le nombre de sources multiplexées devient grand : Notons $\lambda = \mathbb{E}[X]$ et $\sigma^2 = Var[X]$. Si l'on construit la v.a.

$$Z_N = \frac{\sqrt{N}(M_N - \lambda)}{\sigma} \tag{A.9}$$

alors

$$\begin{aligned} \mathbb{P}\left(M_N \geq c\right) &= \mathbb{P}\left[Z_N > \frac{(c-\lambda)\sqrt{N}}{\sigma}\right] \\ Z_N &\to \mathcal{N}(0,1) \\ \lim_{N\to\infty}\mathbb{P}[a < Z_N < b] &= \int_a^b \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx \end{aligned} \tag{A.10}$$

Remarque : Typiquement, le théorème central limite est utilisé pour des valeurs finies de N. Pour "N assez grand" une somme de N variables aléatoires indépendantes suit approximativement une loi normale, dont l'espérance et la variance sont respectivement la somme des espérances et la somme des variances des lois des variables sommées. Pour une précision désirée, à partir de quelle valeur N est-il "assez grand"?. Cela dépend de la loi de X.

$$P[S_N > Nc] \approx \frac{1}{\sqrt{2\pi N\sigma^2}}\int_{Nc}^{\infty} e^{-\frac{(x-N\lambda)^2}{2(N\sigma^2)}}\,dx \tag{A.11}$$

Commentaires

1. Le T.C.L. nous donne un résultat limite.
2. Quelle est l'évolution avec N et la vitesse de convergence ?.
3. Quelle est l'erreur commise avec l'approximation gaussienne ?.
4. Quelle est la dépendance avec la loi de la v.a. pour des v.a. génériques ?

Considérons un cas particulier, $X \sim \mathcal{N}(0,1)$, alors

$$\mathbb{P}[|M_N| \geq c] = 1 - \frac{1}{\sqrt{2\pi}} \int_{-c\sqrt{N}}^{c\sqrt{N}} e^{-x^2/2} dx \quad X \sim \mathcal{N}(0,1) \tag{A.12}$$

On peut montrer que[2]

$$\frac{1}{N} \log \mathbb{P}[|M_N| \geq c] \to -\frac{c^2}{2} \quad , N \to \infty \tag{A.13}$$

et, pour N "assez grand", en dériver des équivalents logarithmiques :

$$\mathbb{P}\left[|M_N| \geq c\right] \approx e^{-N\frac{c^2}{2}} \tag{A.14}$$

Intérêt des Grandes Déviations

Peut-on trouver des résultats similaires à

$$\frac{1}{N} \log \mathbb{P}\left[|M_N| \geq c\right] \to -\frac{c^2}{2} \quad , N \to \infty \tag{A.15}$$

pour des distributions arbitraires de X ? c'est-à-dire, des relations de la forme :

$$\frac{1}{N} \log \mathbb{P}\left[|M_N| \geq c\right] \to -I_X(c) \quad , N \to \infty \tag{A.16}$$

Remarque : Une telle relation revient à dire que la queue de distribution de la moyenne empirique d'une suite de v.a. i.i.d. décroît exponentiellement avec une fonction de taux $I_X(c)$ dépendante de la loi de X. Autrement dit, le logarithme de la probabilité de pertes décroît linéairement avec le nombre de sources multiplexées, avec un coefficient dépendant de la loi de X. Nous verrons dans les sections suivantes comment le théorème de Cramer répond à cette question.

[2] Dans l'intégralité du document, log dénote le logarithme en base e.

A.3. Eléments de la théorie des Grandes Déviations

A.3.1. Quelques définitions et transformées

A.3.1.1. Inégalité de Bienaymé-Chebytchev

Proposition A.3.1 (Inégalité de Bienaymé Chebytchev). *Soit $f : \mathbb{R}^+ \to \mathbb{R}^+$, une fonction mesurable et croissante. Soit X une v.a. à valeurs dans $\mathbb{R}^+$, alors :*

$$P(X > \epsilon) \leq \frac{\mathbb{E}\left[f(X)\right]}{f(\epsilon)} \tag{A.17}$$

Démonstration.

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_\Omega f(X(\omega))\mathbb{P}(d\omega) = \int_{\{X(\omega)>c\}} f(X(\omega))\mathbb{P}(d\omega) + \int_{\{X(\omega)\leq c\}} f(X(\omega))\mathbb{P}(d\omega) \\ &\geq \int_{\{X(\omega)>c\}} f(c)\mathbb{P}(d\omega) + \int_{\{X(\omega)\leq c\}} f(X(\omega))\mathbb{P}(d\omega) \geq f(c) \quad \int_{\{X(\omega)>c\}} \mathbb{P}(dw) = f(c) \quad \mathbb{P}[X > c]\end{aligned}$$

□

A.3.1.2. La Borne de Chernov

La borne de Chernov est un cas particulier de l'inégalité précédente dans laquelle on considère la famille de fonctions exponentielles, paramétrées par $\theta \geq 0$.

$$\begin{aligned}P(X > \epsilon) &\leq \frac{\mathbb{E}\left[e^{(\theta X)}\right]}{e^{\theta\epsilon}} \\ P(X > C) &\leq \mathbb{E}\left[e^{(\theta X)}\right] e^{-\theta C}, \forall\theta \geq 0\end{aligned} \tag{A.18}$$

Bien sûr, la finesse de la borne de Chernov dépend du paramètre θ. En considérant par exemple $\theta = 0$ on retrouve l'inégalité évidente $P(X > x) \leq 1$. Cette borne etant une borne supérieure, on peut s'intéresser à en déterminer la borne la plus fine :

$$P(X > C) \leq \inf_{\theta\geq 0} \mathbb{E}\left[e^{(\theta(X-C))}\right] \tag{A.19}$$

Dans les sections suivantes on détaillera le rapport entre l'équation A.19, le théorème de Cramer et le Principes de Grandes Déviations. Notons qu'une simple manipulation des termes nous permet d'écrire (en utilisant les propriétés des fonctions $\log(x)$ et $\exp(x)$) :

$$\begin{aligned}P(X > C) &\leq \inf_{\theta\geq 0} \mathbb{E}\left[e^{(\theta X)}\right] e^{-\theta C} = \\ &= e^{-\sup_{\theta\geq 0}\left\{\theta C - \log \mathbb{E}\left[e^{(\theta X)}\right]\right\}}\end{aligned} \tag{A.20}$$

A.3.1.3. Fonction Génératrice Logarithmique (ou Log-Laplace)

Définition A.3.1 (Tranformée de Log Laplace). Soit $X(0,t)$ un processus stochastique cumulatif (processus d'accroissements ou par abus du langange, si les incréments du processus sont des

entiers positifs, processus de comptage). Les transformées de (Log) Laplace [3] sont définies par :

$$\begin{aligned} &X(0,t) \overset{\mathcal{T}.\mathcal{L}.}{\rightarrow} L_X(\theta) \\ &X(0,t) \overset{\mathcal{T}.\mathcal{L}\mathcal{L}.}{\rightarrow} \Lambda_X(\theta) \\ &L_X(\theta) \triangleq \mathbb{E}\left\{e^{\theta X(0,t]}\right\} \\ &\Lambda_X(\theta) \triangleq \log L_X(\theta) = \log \mathbb{E}\left\{e^{\theta X(0,t]}\right\} \end{aligned} \tag{A.21}$$

Lemme A.3.1 (Convexité de la transformée Log Laplace). *La transformée Log Laplace est convexe.*

Démonstration. D'après l'inégalité de Holder

$$\begin{aligned} \Lambda_X(\alpha\theta_1 + (1-\alpha)\theta_2) &= \log \mathbb{E}\left\{\left(e^{\theta_1 X(0,t]}\right)^{\alpha}\left(e^{\theta_2 X(0,t]}\right)^{(1-\alpha)}\right\} \\ &\leq \log\left\{\mathbb{E}[e^{\theta_1 X(0,t]}]^{\alpha}\mathbb{E}[e^{\theta_2 X(0,t]}]^{(1-\alpha)}\right\} \\ &= \alpha\Lambda_X(\theta_1) + (1-\alpha)\Lambda_X(\theta_2) \end{aligned} \tag{A.22}$$

□

Remarque A.3.1 *(Bande Passante Effective (ou Equivalente)).* Kelly [53] et d'autres auteurs ont popularisé la notion de bande passante effective, directement en rapport avec la transformée, (cf. chapitre 5).

$$\alpha(s,t) = \frac{1}{st}\log \mathbb{E}\left\{e^{sX(0,t)}\right\} \tag{A.23}$$

A.3.1.4. Transformée de Fenchel-Legendre (Transformée Convexe)

Définition A.3.2 (Transformée Convexe). Soit une fonction $\Lambda(\theta) : \mathbb{R}^d \rightarrow \mathbb{R}\bigcup\{+\infty\}, X$ avec $\theta \in \mathbb{R}^d$. La transformée de Fenchel-Legendre de $\Lambda(\theta)$ est définie de la façon suivante :

$$\begin{aligned} &\Lambda(\theta) \overset{\mathcal{T}.\mathcal{F}-\mathcal{L}.}{\rightarrow} \Lambda^*(x) \\ &\Lambda^*(x) = \sup_{\theta\in\mathbb{R}^d}\{<\theta, x> -\Lambda(\theta)\} \end{aligned} \tag{A.24}$$

Propriétés de la Transformé de Fenchel-Legendre

– (a) La transformée de Fenchel-Legendre est convexe.

Lemme A.3.2 (Convexité).

Démonstration. (cas $\mathbb{R}$)

$$\begin{aligned} \alpha\Lambda^*(x_1) + (1-\alpha)\Lambda^*(x_2) &= \sup_{\theta\in\mathbb{R}}\{\alpha\theta x_1 - \alpha\Lambda(\theta)\} + \sup_{\theta\in\mathbb{R}}\{(1-\alpha)\theta x_2 - (1-\alpha)\Lambda(\theta)\} \\ &\geq \sup_{\theta\in\mathbb{R}}\{(\alpha x_1 + (1-\alpha)x_2)\theta - \Lambda(\theta)\} \\ &= \Lambda^*(\alpha x_1 + (1-\alpha)x_2) \end{aligned} \tag{A.25}$$

□

[3] (ang. Logarithmic Moment Generating Function ou Cumulant Generating Function)

– (b) La transformée de Fenchel Legendre est semi continue inférieurement ($\liminf y \to x \quad \Lambda^*(y) \geq \Lambda^*(x)$).

Remarque A.3.2 *(Application).* Dans le cas où $\Lambda(\theta)$ est la transformée de LogLaplace d'une v.a X, $X \in \mathbb{R}$, voire d'un processus stochastique cumulatif $X(0,t)$, dans $\mathbb{R}$, nous avons :

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \log \mathbb{E} \left\{ e^{\theta X(0,t)} \right\} \right\} \tag{A.26}$$

Interprétation Géométrique de la Transformée de Fenchel-Legendre La figure A.2 illustre l'interprétation géométrique de la transformée de Fenchel Legendre. La valeur de θ qui réalise le maximum correspond au point où la droite tangente à la courbe $\Lambda(\theta)$ a la même pente x que la droite définie par la fonction $f(\theta) = \theta x$. Ensuite, on mesure la différence entre la droite et la fonction.

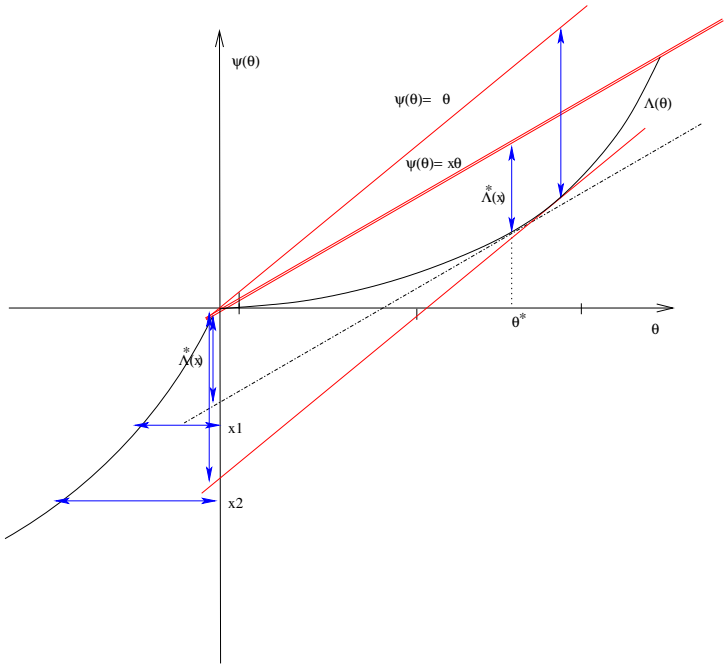


Fig. A.2: Interprétation Géométrique de la Transformée de Fenchel Legendre

A.3.2. Théorème de Cramer

Le théorème de Cramer est utile lorsque l'on souhaite déterminer une estimation de la probabilité que sous un certain régime asymptotique, la moyenne empirique d'une suite de variables i.i.d. prenne une valeur dans un intervalle particulier.

Théorème A.3.1 (Théorème de Cramer). *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d prenant leurs valeurs dans $\mathbb{R}$. Le théorème de Cramer établit que pour tout ensemble ouvert G, et*

pour tout ensemble fermé, F, la moyenne empirique M_n définie par $n^{-1}\sum_{i=1}^{n} X_i$ vérifie :

$$\begin{aligned} \limsup_{n\to\infty} \quad & \frac{1}{n}\log \mathbb{P}\left[M_n \in F\right] \leq -\inf_{x\in F} \Lambda^*(x) \\ \liminf_{n\to\infty} \quad & \frac{1}{n}\log \mathbb{P}\left[M_n \in G\right] \geq -\inf_{x\in G} \Lambda^*(x) \end{aligned} \tag{A.27}$$

où Λ^* est la transformée de Fenchel Legendre de la Log Laplace de la variable générique X.

Un cas particulier intéressant est $F = [a, +\infty)$, ou $a > \mathbb{E}[X]$. La limite supérieure est donnée par l'expression suivante :

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left[M_n \in F\right] &\leq -\inf_{x\in F} \Lambda^*(x) \\ \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left[M_n \geq a\right] &\leq -\inf_{x\geq a} \Lambda^*(x) \\ &\leq -\inf_{x\geq a}\sup_{\theta\in\mathbb{R}} \left\{\theta x - \Lambda_X(\theta)\right\} \\ &\leq -\sup_{\theta\geq 0} \left\{\theta a - \Lambda_X(\theta)\right\} \end{aligned}$$

Notons que grâce à la convexité de la transformée de Fenchel Legendre de la transformée de Log Laplace (fonction de taux) d'une variable aléatoire, le minimum sur l'intervalle $[a, +\infty)$ est réalisé en $x = a$, et que le terme $\theta x - \Lambda_X(\theta)$ étant concave (unimodal) par rapport à θ, le supremum par rapport à θ est restreint à des valeurs de $\theta \geq 0$. Remarquons le rapport avec la borne de Chernov présentée précédemment.

A.3.3. Principe de Grandes Déviations

Définition A.3.3 (Principe de Grandes Déviations (PGD)). Une suite de mesures de probabilité μ_n sur un espace métrique $\mathcal{X}$ suit un Principe de Grandes Déviations de vitesse $\delta_n \to +\infty$ et de fonction de taux $\Lambda^* : \mathcal{X} \to [0, \infty]$ si pour tout ensemble ouvert G et tout ensemble fermé F tels que $F, G \subset \mathcal{X}$

$$\begin{aligned} -\inf_{x\in G} \Lambda^*(x) &\leq \liminf_{n\to\infty} \frac{1}{\delta_n}\log \mu_n(G) \\ &\leq \limsup_{n\to\infty} \frac{1}{\delta_n}\log \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x) \end{aligned} \tag{A.28}$$

Remarque A.3.3 *(Définition équivalente).* Soit Y_n une suite de variables aléatoires dans un espace de Hausdorff (complet, séparable, métrique) E, (typiquement, $\mathbb{R}$). Soit $I : E \to \mathbb{R}^+ \bigcup\{+\infty\}$. Soit $F \in E$ un ensemble fermé, Soit $G \in E$ un ensemble ouvert. On dit que la suite Y_n suit un *Principe de Grandes Déviations (ang. Large Deviation Principle)* de fonction de taux I si et seulement si :

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left[Y_n \in F\right] \leq -\inf_{y\in F} I(y) \\ \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left[Y_n \in G\right] \geq -\inf_{y\in G} I(y) \end{aligned} \tag{A.29}$$

Exemple A.3.1 *(Le théorème de Cramer comme PGD).* Le théorème de Cramer peut être exprimé sous forme de PGD avec : $Y_n = \frac{\sum_{i=1}^{n} x_i}{n}, I = \Lambda^*$.

Définition A.3.4 (Ensembles de Continuité). Soit C un ensemble ouvert et $\bar{C}$ son fermé. Si le minimum de I est atteint en un point vérifiant :

$$\inf_{y \in C} I(y) = \inf_{y \in \bar{C}} I(y)$$

alors on dit que C est un ensemble de continuité.

Si l'ensemble considéré est un ensemble de continuité, alors le PGD prend une expression plus simple :

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[Y_n \in C\right] = - \inf_{y \in C} I(y)$$
$$\mathbb{P}(Y_n \in C) \approx e^{-n \inf_{y \in C} I(y)}$$

A.3.4. Le théorème de Gartner-Ellis

Le théorème de Gartner-Ellis [36][27] est une extension directe du théorème de Cramer pour les suites de variables aléatoires non nécessairement i.i.d. Formalisons cette idée : si, pour une suite de v.a. Y_n , la limite suivante (limite de Gartner-Ellis)

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}\left\{e^{\theta Y_n}\right\} = \Lambda(\theta)$$

existe et est finie pour toutes les valeurs $\theta \in \mathbb{R}$ et cette limite est différentiable dans son domaine, alors Y_n/n suit un principe de grandes déviations dont la fonction de taux est donnée par la transformée convexe de la limite de Gartner - Ellis de Y_n. Dans le cadre du télétrafic, considérons l'exemple suivant : soit $A(s,t]$ un processus d'arrivées (par exemple $A(s,t] = \sum_{n \in \mathbb{Z}} \mathbb{1}_{\{T_n \in (s,t]\}} \sigma_n$), et définissons $Z_s \triangleq A(-s,0]$.

A.3.4.1. Exemple : Processus de Poisson

Dans cet exemple, considérons d'abord un processus de Poisson, $A \sim Poisson(\lambda), \sigma_n = 1$ (arrivés simples). Nous avons donc :

$$\begin{aligned}
\mathbb{E}\left\{e^{\theta A(-s,0]}\right\} &= \sum_{n=0}^{\infty} e^{\theta n} \mathbb{P}\left(A(-s,0] = n\right) \\
&= \sum_{n=0}^{\infty} e^{\theta n} \frac{(\lambda(-s,0))^n}{n!} e^{-\lambda(-s,0)} \\
&= e^{-\lambda s} \sum_{n=0}^{\infty} e^{\theta n} \frac{(\lambda s)^n}{n!} \\
&= e^{-\lambda s} e^{\lambda s e^{\theta}} = e^{\lambda s \left(e^{\theta} - 1\right)}
\end{aligned}$$

la limite de Gartner Ellis est donnée par :

$$\Lambda(\theta) = \lim_{s\to\infty} \frac{1}{s} \log \mathbb{E}\left\{e^{\theta A(-s,0]}\right\} = \lambda(e^{\theta} - 1)$$

Hypothèse vérifiée, alors $A(-s, 0]/s$ vérifie un PGD dont la fonction de taux est donnée par :

$$\begin{aligned} h(x) &= \sup_{\theta} \left\{\theta x - \lambda(e^{\theta} - 1)\right\} \\ &= \begin{cases} x \log(\frac{x}{\lambda}) - (x - \lambda) & x \geq 0 \\ +\infty & x < 0 \end{cases} \end{aligned}$$

A.3.4.2. Exemple : Processus de Renouvellement [69]

Hypothèses de départ : soit N un processus ponctuel, $T_1, T_2, \ldots, T_n$ avec

$$\begin{aligned} \sigma_n &\triangleq 1 \qquad \mathbb{E}[e^{\theta T_1}] < \infty \quad \forall \theta > 0 \\ \lambda &\triangleq \frac{1}{\mathbb{E}_N^o[T_1]} \end{aligned}$$

Caractérisation de $\mathbb{E}[e^{\theta T_n}]$ et calcul de la limite de Gartner-Ellis : Notons

$$\mathbb{P}\left(N(0,s) \geq a\right) = \mathbb{P}\left(\frac{N(-s,0)}{s}\right) = \mathbb{P}\left(T_{[as]} < s\right)$$

$$e^{\theta T_n} \quad o \quad \theta_t = e^{\theta(T_n - t)}$$

D'après la formule d'inversion,

$$\begin{aligned} \mathbb{E}[f] &= \lambda \mathbb{E}_N^o \left[\int_0^{T_1} f \quad o \quad \theta_t dt\right] \\ \mathbb{E}[e^{\theta T_n}] &= \lambda \mathbb{E}_N^o \left[\int_0^{T_1} e^{\theta T_n} \quad o \quad \theta_t dt\right] \end{aligned}$$

$$\begin{aligned} \mathbb{E}[e^{\theta T_n}] &= \frac{1}{\mathbb{E}_N^o[T_1]} \mathbb{E}_N^o \left[\int_0^{T_1} e^{\theta(T_n - t)} dt\right] = \frac{1}{\mathbb{E}_N^o[T_1]} \mathbb{E}_N^o \left[e^{\theta T_n} \int_0^{T_1} e^{-\theta t} dt\right] \\ &= \frac{1}{\mathbb{E}_N^o[T_1]} \mathbb{E}_N^o \left[e^{\theta T_n} \frac{1 - e^{-\theta T_1}}{\theta}\right] = \frac{1}{\mathbb{E}_N^o[T_1]\theta} \left(\mathbb{E}_N^o[e^{\theta T_n}] - \mathbb{E}_N^o[e^{\theta(T_n - T_1)}]\right) \end{aligned}$$

Notons que

$$\mathbb{E}_N^o[e^{\theta T_n}] = \mathbb{E}_N^o\left[e^{\theta \sum_{i=1}^n (T_i - T_{i-1})}\right] = \prod_{i=1}^n \mathbb{E}_N^o[e^{\theta T_1}] = \mathbb{E}_N^o[e^{\theta T_1}]^n$$

Alors

$$\mathbb{E}[e^{\theta T_n}] = \frac{1}{\mathbb{E}_N^o[T_1]\theta} \left(\mathbb{E}_N^o[e^{\theta T_1}]^n - \mathbb{E}_N^o[e^{\theta(T_n - T_1)}]\right)$$

Finallement

$$\frac{1}{n}\log\mathbb{E}[e^{\theta T_n}] \to \log\mathbb{E}_N^o[e^{\theta T_1}] \triangleq \Lambda_N^o(\theta)$$
$$\text{T.F-L.}\,\{\Lambda_N^o(\theta)\} = \Lambda_N^{o\,*}(x) \triangleq h(x)$$

Principes de grandes déviations dérivés Le théorème de Gartner-Ellis nous permet d'affirmer que la suite $n^{-1}T_n$ suit un PGD de fonction de taux $h(x)$ et que

$$\frac{1}{s}\log\mathbb{P}\left(N(-s,0)\right) \to -ah\left(\frac{1}{a}\right)$$

A.4. Grandes Déviations Trajectorielles

Les principes des grandes déviations tels qu'on les a présentés permettent le calcul du comportement asymptotique de suites de v.a., notamment la queue de distribution des moyennes empiriques. Il est parfois intéressant de caractériser le comportement d'une trajectoire d'un processus stochastique. Ceci est l'objectif de la théorie des grandes déviations trajectorielles (*ang. Sample Path Large Deviations*)

Soit D l'espace des fonctions $\mathbb{R} \to \mathbb{R}$, continues à droite avec limite à gauche (c.à.d.l.à.g), muni de la topologie de la convergence uniforme sur les ensembles compacts, c'est à dire : soit $f \in D$, $f_n \in D$,

$$f_n \to f \iff \forall T > 0,\ \lim_{n\to\infty} \sup_{|t|<T} |f_n(t) - f(t)| \to 0$$

Soit $A(t)$ définie :

$$A(t) = \begin{cases} X(0,t] & \text{pour } t \geq 0 \\ -X(t,0] & \text{pour } t < 0 \end{cases} \tag{A.30}$$

On peut noter $A(s,t) = A(t) - A(s) \quad \forall s < t$. On fait l'hypothèse que $A(t) \in D, A(t)$ est un processus croissant, à accroissements positifs. Soit $A^{(N)}(t) = \frac{A(Nt)}{N}$ (normalisation en temps et en space). Alors $\exists h : \mathbb{R} \to \mathbb{R}^+ \bigcup\{+\infty\}$, function strictement convexe, admettant un unique minimum en $\lambda > 0$, avec $h(\lambda) = 0$ telle que la suite de processus $A^{(N)}(t)$satisfait un PGD dans D de fonction de taux $I : D \to \mathbb{R}^+ \bigcup\{+\infty\}$, définie

$$I(a) = \begin{cases} \int_{-\infty}^{+\infty} h(\dot{a})dt & \text{si a(t) est absolument continue et a(0) = 0} \\ +\infty & \text{sinon.} \end{cases} \tag{A.31}$$

intervient dans un PGD de la forme (cf. [13], Mogulskii theorem)

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{N}\log\mathbb{P}\left[A^{(N)} \in F\right] &\leq -\inf_{f\in F} I(f) \\ \liminf_{N\to\infty} \frac{1}{N}\log\mathbb{P}\left[A^{(N)} \in G\right] &\geq -\inf_{f\in G} I(f) \end{aligned} \tag{A.32}$$

Par exemple, si $A(t)$ est un processus de renouvellement, $\{T_n\}$, avec $\exists\theta > 0$ tel que $\mathbb{E}_N^o\left(e^{\theta T_1}\right) < \infty$ et $\mathbb{P}^o\left(T_1 > 0\right) = 1$, alors l'hypothèse est vérifiée, avec :

$$\begin{aligned} h(x) &= x\Lambda^*\left(\frac{1}{x}\right) \\ \Lambda^*\left(x\right) &= \sup_\theta\left(\theta x - \Lambda(\theta)\right) \\ \Lambda\left(\theta\right) &= \log \mathbb{E}_N^o\left(e^{\theta T_1}\right) \end{aligned} \tag{A.33}$$

A.5. Le principe de Contraction

Le principe de contraction permet d'obtenir des PGDs (cf. section A.3.3) à partir d'un PGD donné. Le principe de contraction s'avère un outil très puissant pour calculer des fonctions de taux pour des suites de variables aléatoires obtenues comme fonction d'une suite de variables aléatoires dont on connaît la fonction de taux.

Théorème A.5.1 (Principe de Contraction [23]). *Soient E et F deux espaces topologiques de Hansdorff, soit X_n une suite de variables aléatoires à valeurs dans E vérifiant un PGD de fonction de taux $I(x)$. On fait l'hypothèse que les ensembles $\mathcal{X}_A \triangleq \{x : I(x) \leq A, \forall A > 0\}$ sont compacts. Soit $\phi : E \to F$ une fonction continue sur chaque ensemble $\mathcal{X}_A$. Alors, la suite $Y_n = \phi(X_n)$ à valeurs dans F vérifie un PGD de fonction de taux $J(y) = \inf_{y:\phi(y)=x} I(x)$.*

A.6. Applications

A.6.1. Multiplexage sans Buffer avec des sources de Poisson

Dans ce premier cas, nous allons voir comment utiliser la borne de Chernov pour le dimensionnement et le contrôle d'admission dans un système à temps discret sans buffer, de capacité C. A chaque slot de temps, N sources i.i.d. produisent une quantité de travail (e.g. en nombre de paquets, bits, etc.) modélisée par une v.a. de Poisson (choix justifié du fait de la simplicité de sa bande passante effective et à titre d'exemple). Nous notons X_i le travail produit par une source pendant un slot de temps et $X_T = \sum_{i=1}^N X_i$ le travail total produit pour l'agrégat pendant le même slot de temps (X_T représente les arrivées totales dans le système dans un slot de temps). Nous avons donc :

$$\begin{aligned} \mathbb{E}[e^{\theta X_i}] &= e^{\lambda(e^\theta - 1)} \\ \mathbb{E}[e^{\theta X_T}] &= e^{N\lambda(e^\theta - 1)} \\ \log \mathbb{E}[e^{\theta X_i}] &= \Lambda_i(\theta) = \lambda(e^\theta - 1) \\ \log \mathbb{E}[e^{\theta X_T}] &= \Lambda_T(\theta) = N\lambda(e^\theta - 1) \end{aligned} \tag{A.34}$$

Evaluation : En appliquant la borne de Chernov, nous obtenons :

$$\begin{aligned} \mathbb{P}(X_T > C) &\leq \inf_{\theta \geq 0} e^{N\lambda(e^\theta - 1) - \theta C} \\ \mathbb{P}(X_T > C) &\leq e^{\inf_{\theta \geq 0} N\lambda(e^\theta - 1) - \theta C} \end{aligned} \tag{A.35}$$

avec

$$J(\theta) \triangleq N\lambda(e^{\theta} - 1) - \theta C \tag{A.36}$$

La valeur de θ^* qui minimise $J(\theta)$ (réalisant la borne la plus fine) est donnée par :

$$\theta^* = \log \frac{C}{N\lambda} \tag{A.37}$$

Nous avons donc :

$$\mathbb{P}(X_T > C) \leq e^{C - N\lambda - C\log\left(\frac{C}{N\lambda}\right)} \tag{A.38}$$

Cette expression permet de calculer des paramètres d'intérêt tels que le nombre maximum de sources à admettre pour ne pas dépasser la capacité avec une probalité donéée, etc.

Dimensionnement du système : Si nous fixons $\lambda = 1$, la figure A.3 illustre la probabilité $\mathbb{P}(X_T > C)$ en fonction de le nombre de sources multiplexés N pour des valeurs de $C = 100, 110, 120, 130$. Avec ce type d'analyses, il est possible de calculer la capacité à déployer afin de garantir une qualité de service souhaitée.

Contrôle d'admission : également, la figure A.4 montre le nombre de sources que le système peut accepter afin de garantir une qualité de service donnée.

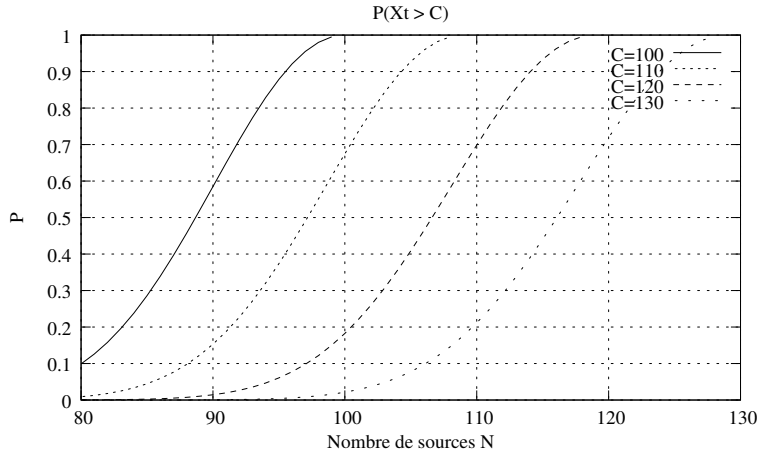


Fig. A.3: Dimensionnement du système. Application de la borne de Chernov.

A.6.2. Duffield-O'Connell : L'asymptotique de grand buffer

Avant de présenter l'asymptotique de grand buffer, nous faisons une synthèse d'un résultat connu sous le nom de borne de Kingman. Nous verrons ensuite comment l'asymptotiquede grand buffer généralise les conditions d'applicabilité de la borne de Kingman.

Borne de Kingman pour les files GI/GI/1

Soit $(\tau, \sigma) = (\tau_n, \sigma_n)$ une suite modélisant les temps d'inter-arrivées et les temps de service d'une file GI/GI/1 sous les hypothèses suivantes :

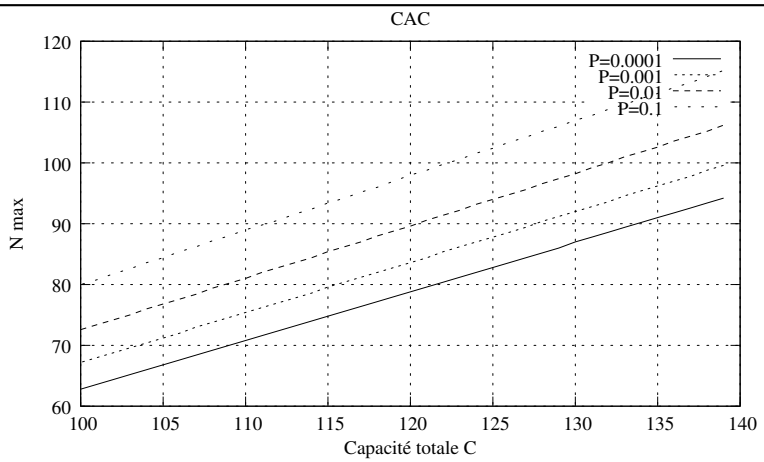


Fig. A.4: Contrôle d'admission. Application de la borne de Chernov.

1. $\exists\theta > 0 \quad | \quad \mathbb{E}(e^{\theta(\sigma-\tau)}) = 1$
2. $\mathbb{E}[\sigma] < \mathbb{E}[\tau]$ (la file est stable).

Remarque : Sous les hypothèses mentionnées, la convexité de la transformée de Log Laplace $\log \mathbb{E}[e^{\theta(\sigma-\tau)}]$, et le fait que celle-ci vaut zéro en $\theta = 0$ implique que la valeur minimale est atteinte au point $\theta_{\min}$ (la valeur qui annule la dérivée de la Loglaplace), car $\log \mathbb{E}\left[e^{\theta_{\min}(\sigma-\tau)}\right] < 0$ et que la pente en zéro est négative $\frac{\partial \log \mathbb{E}(e^{\theta(\sigma-\tau)})}{\partial\theta}\Big|\theta = 0 = \mathbb{E}(\sigma - \tau) < 0$. Voir figure A.5.

Soit W_n le temps d'attente du client n (rapellons qu'il s'agit d'une file d'attente à discipline FIFO, et à conservation de travail). Suppons que $W_1 = 0$. D'après l'équation de Lindley et le théorème de Loynes,

$$W_{n+1} = \max\left(W_n + \sigma_n - \tau_n, 0\right)$$
$$\lim_{n\to\infty} \mathbb{P}(W_n \leq x) = \mathbb{P}(W \leq x)$$

Alors, la borne de Kingman [54] donne :

$$P(W_n \geq x) \leq e^{-\theta^* x} \quad \forall x > 0, n \geq 1$$
$$P(W \geq x) \leq e^{-\theta^* x}$$

Liu et Nain [59] proposent des extensions à la borne de Kingman pour des cas plus génériques.

Nous désirons caractériser la queue de distribution du processus "travail cumulé" dans une file d'attente dans de conditions moins restrictives que les hypothèses de Kingman. Selon que la taille du buffer est finie ou non, nous notons W^B la distribution stationnaire du travail cumulé dans une file à buffer de taille B et W pour la distribution stationnaire du travail cumulé dans une file d'attente à buffer de taille infinie. Rappelons la relation suivante :

$$P(W^B = B) \leq P(W \geq B)$$

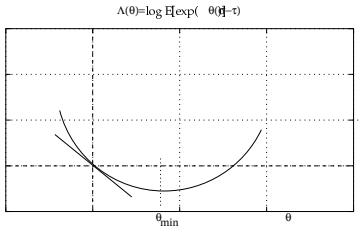


Fig. A.5: Borne de de Kingman : Interprétation géométrique

L'asymptotique de grand buffer, obtenue par Duffield et O'Connell [25] nous permet de caractériser la queue de distribution du travail cumulé dans une file d'attente de capacité C, au moyen d'un principe de grandes déviations sous un certain nombre d'hypothèses, notamment sous l'existance d'un principe de grandes déviations du processus d'arrivées. Soit $A(s,t]$ un processus d'arrivés. Soit Z_s le travail produit par une source pendant $(-s,0)$. Supposons que le débit moyen empirique, $\frac{Z_s}{s}$ suit un principe de grandes déviations avec fonction de taux $\mathrm{h}(x)$. On note l'intensité du processus $\lambda < C$. On fait l'hypothèse que $\inf_{u\geq C} \mathrm{h}(u) > 0$, et on note $W = W(0)$ le travail stationnaire en 0. Alors,

$$\limsup \frac{1}{x} \log \mathbb{P}(W \geq x) \leq - \inf_{\tau>0} \tau \inf_{y\geq 0} \mathrm{h}(C + y + \frac{1}{\tau}) = - \inf_{\tau>0} \tau \inf_{u\geq C+1/\tau} \mathrm{h}(u)$$
$$\liminf \frac{1}{x} \log \mathbb{P}(W > x) \geq - \inf_{\tau>0} \tau \inf_{y>0} \mathrm{h}(C + y + \frac{1}{\tau}) = - \inf_{\tau>0} \tau \inf_{u> C+1/\tau} \mathrm{h}(u)$$

Asymptotique de Grand Buffer avec Fonction de Taux Convexe : Si la fonction de taux h est convexe, pour une valeur de τ fixée, l'infimum par rapport à y implique $y \to 0$.

$$\begin{aligned}\limsup \frac{1}{x} \log \mathbb{P}(W \geq x) &\leq - \inf_{\tau>0} \tau \inf_{y\geq 0} \mathrm{h}(C + y + \frac{1}{\tau}) \\ &= - \inf_{\tau>0} \tau \mathrm{h}(C + \frac{1}{\tau}) \\ &= - \inf_{\sigma>0} \frac{\mathrm{h}(C+\sigma)}{\sigma}\end{aligned}$$

si l'on définit

$$\theta^* = \inf_{\sigma>0} \frac{\mathrm{h}(C+\sigma)}{\sigma}$$

alors

$$\frac{1}{x} \log \mathbb{P}(W \geq x) \to -\theta^*$$
$$\mathbb{P}(W \geq x) \approx e^{-x\theta^*}$$

Les fonctions $\Lambda(\theta)$ et $h(x)$ sont *convex conjugates,* et on peut montrer que

$$\Lambda(\theta^*) = \sup_x \{\theta^* x - h(x)\} = C\theta^*$$

Interprétation : Le logarithme de la probabilité de débordement décroît linéairement avec la taille du buffer (pour des tailles assez grandes), avec un coefficient θ^* solution de $\frac{\Lambda(\theta)}{\theta} = C$, où $\Lambda(\theta)$ est la limite de Gartner-Ellis du processus d'arrivées. Autrement dit, (cf. figure A.6), la valeur de θ^* est la pente de la droite tangente à la transformée de Fenchel Legendre de la limite de Gartner Ellis citée, qui passe par $(C, 0)$, c.-à.-d. $y(x) = \theta^*(C - x)$.

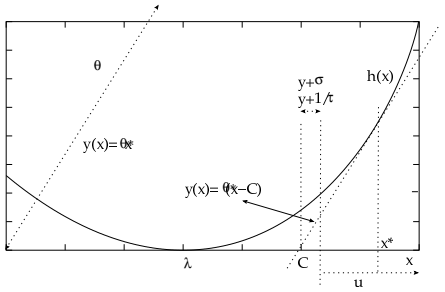


Fig. A.6: Illustration Graphique de l'asymptotique de grand buffer

Contrôle d'admission

Le résultat de Duffield-O'Connell a souvent été utilisé pour définir un critère de contrôle d'admission : Supposons que le trafic agrégé constituant le processus d'arrivées d'une file d'attente est composé d'un certain nombre N de sources indépendantes. Le processus d'arrivées associé à chaque source est noté $A_i(s,t)$ et l'agrégat $A(s,t) = \sum A_i(s,t)$. Nous faisons aussi l'hypothèse que chaque source vérifie les conditions de Gartner-Ellis, (voir section A.3.4), c'est à dire, la limite $\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}(e^{\theta A_i(-t,0)})$ existe et est notée $\Lambda_i(\theta)$. Soit $\Lambda(\theta) = \sum \Lambda_i(\theta)$, la limite de Gartner-Ellis du processus $A(s,t)$. D'après l'asymptotique de grand buffer :

$$\mathbb{P}(W \geq x) \approx e^{-x\theta^*}$$

Autrement dit, un critère de performance basé sur la valeur de $\mathbb{P}(W \geq x) = \gamma$ impose une valeur maximale θ^*.

$$\theta^* = -\frac{1}{x} \log(\gamma)$$

Remarquons que cette relation est vraie pour de grandes valeurs de x. L'ensemble des sources peut être admis en respectant le critère donné si :

$$\frac{\Lambda(\theta^*)}{\theta^*} \leq C$$
$$\sum_{i=1}^{N} \frac{\Lambda_i(\theta^*)}{\theta^*} \leq C$$

Et chaque source se verra attribuer une bande passante effective (additive) égale à :

$$\frac{\Lambda_i(\theta^*)}{\theta^*}$$

A.6.3. L'asymptotique de Grand Nombre d'usagers (Botvich-Duffield, Simonian-Guibert, Weber-Courcoubetis, et Likhanov-Mazumdar)

Considérons une file d'attente de capacité $C = Nc$. Le trafic offert à la file d'attente, $A^{(N)} = \sum_{n=1}^{N} A_n(s,t)$ est la superposition de N sources, i.i.d. où $A_n(s,t]$ est le travail produit par la source n sur l'intervalle de temps $(s,t]$. Soit $W^{(N)}$ le travail cumulé dans la file d'attente. D'après l'équation de Lindley,

$$W^{(N)}(t) = \sup_{s \leq t} \left\{ \sum_{n=1}^{N} A_n(s,t) - C(t-s) \right\} \tag{A.39}$$

La probabilité de débordement d'un seuil ou la borne supérieure de la probabilité de saturation du système à buffer fini est donnée par :

$$\mathbb{P}\left(W^{(N)}(0) \geq Nb\right) \tag{A.40}$$

Sous les hypothèses

1. Chaque A_n a des accroissements ≥ 0, et est stationnaire.
2. $\forall t > 0, \exists \theta > 0 \quad \left|\mathbb{E}[e^{\theta A(0,t]}] < \infty\right.$
3. $\Lambda_t(\theta) = \log \mathbb{E}\left[e^{\theta A(0,t]}\right]$
4. $I_t(x) = \sup_\theta \{\theta x - \Lambda_t(\theta)\}$
5. $\lambda = \mathbb{E}[A(0,1]] < c$
6. $\exists \alpha > 0 | \frac{I_t(ct)}{t^\alpha} \geq \varepsilon > 0, t \to \infty$

alors,

$$\begin{aligned} -J(b^+) &\leq \liminf_{N\to\infty} \quad \frac{1}{N} \log \mathbb{P}\left[W^{(N)} > Nb\right] \\ &\leq \limsup_{N\to\infty} \quad \frac{1}{N} \log \mathbb{P}\left[W^{(N)} \geq Nb\right] \\ &\leq -J(b^-) \end{aligned} \tag{A.41}$$

où

$$J(b) = \inf_{t>0} I_t\left(b + ct\right) \tag{A.42}$$

Cette asymptotique étant un résultat clé de nos travaux et est présentée en détail au chapitre 5, *Outils Mathématiques*.

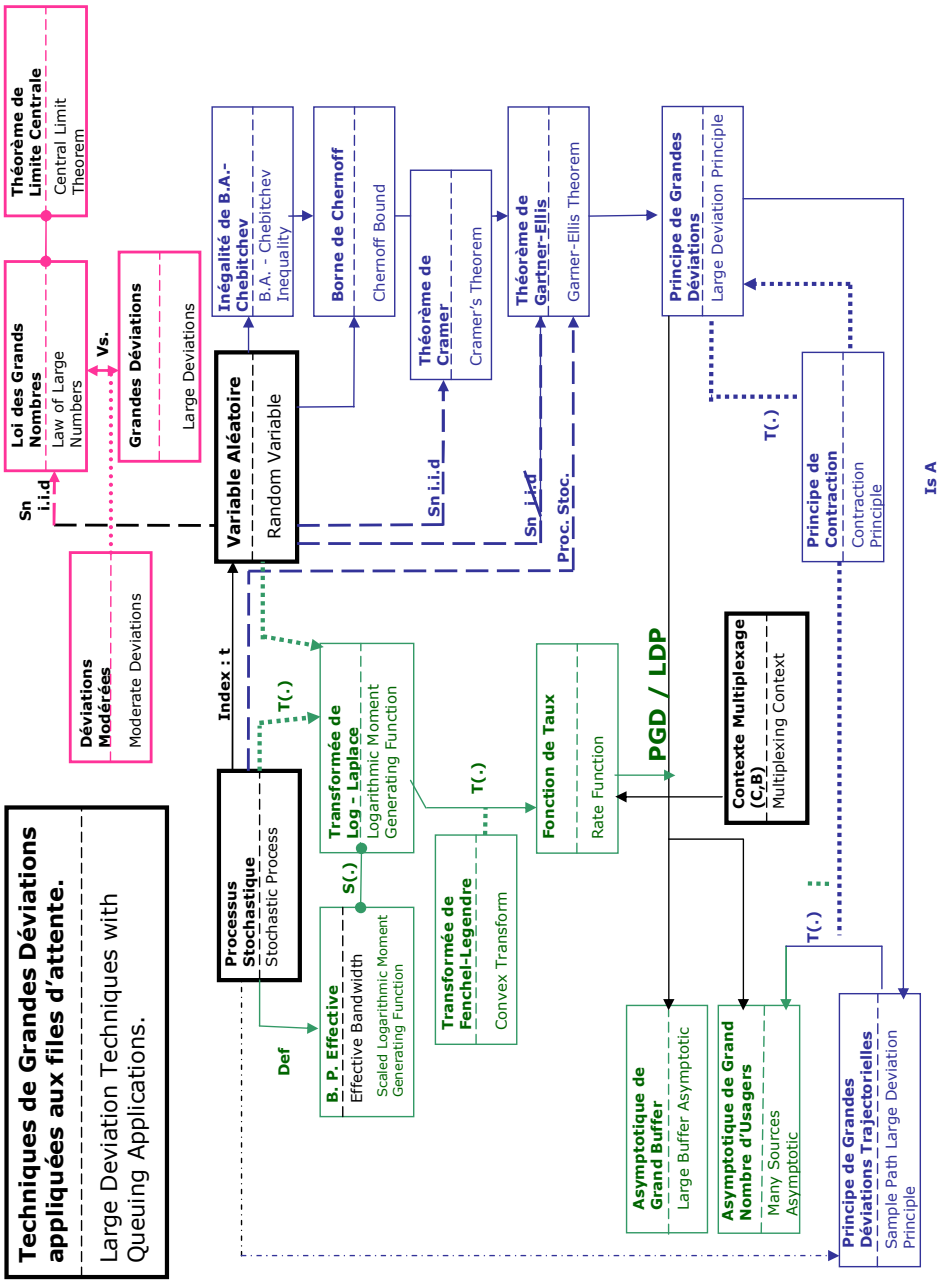


Fig. A.7: Vision d'Ensemble de la théorie des Grandes Déviations

B. Eléments Logiciel

B.1. Motivation

Comme on l'a évoqué dans la conclusion, bien qu'elle soit très générale, la théorie des b.p.e. fait souvent appel à des calculs numériques pour des cas non triviaux. Les simulations et les différents calculs et optimisations numériques des études de cette thèse ont été réalisés en utilisant le langage C++ et un certain nombre de librairies spécialisées dont nous présentons ici les éléments essentiels.

B.2. Librairies utilisés

B.2.1. Standard Template Library (STL)

Les structures de donnés utilisées dans les simulations et calculs numériques des chapitres 6,7,8 et 9 utilisent cette bibliothèque, notamment pour les conteneurs, listes chaînées et algorithmes.

B.2.2. Boost Library

Des membres du comité de standardisation du langage C++ ont formé le projet Boost [1], qui apporte des extensions au langage C++ susceptibles d'être incorporées dans de nouvelles versions du langage. En ce qui nous concerne, nous avons utilisé le module de génération de nombres aléatoires, le module de "Smart Pointers", pour la gestion automatique de la mémoire et le "Boost Graph Library", bibliothèque pour la gestion de graphes. La BGL [2] a été utilisée dans le contexte des travaux de la partie finale de la thèse, en particulier pour les extensions des résultats à des réseaux.

[1] http ://www.boost.org

[2] (c) Boost www.boost.org et Jeremy Siek Indiana University (jsiek@osl.iu.edu). Portions reproduites avec sa permission

C. MPLS et Le système d'exploitation Linux

C.1. MPLS for Linux

MPLS pour Linux est un projet sous forme de logiciel libre destiné à implémenter un ensemble de protocoles de signalisation et le plan d'acheminement *(forwarding plane)* pour le système d'exploitation Linux. C'est un projet très actif, dont le développement a lieu à sourceforge *http ://mpls-linux.sourceforge.net* et *http ://www.sourceforge.net/projects/mpls-linux.*

Aujourd'hui, mpls-linux a déjà développé certaines fonctionnalités du plan usager (data plane) et du plan de contrôle, et peut être intégré dans la série de noyaux Linux 2.4.X (le code source existe sous forme de patch). Il existe de plus une implémentation du protocole LDP, décrit dans [RFC3036].

MPLS pour Linux, est pour l'instant composé de deux modules logiciels :

1. *mpls-linux* Acheminement MPLS basé sur le concept de la commutation d'étiquettes, intégré dans le noyau. Ses principaux atouts sont :
   - Prise en charge des interfaces Ethernet.
   - Prise en charge des interfaces ATM.
   - Prise en charge des interfaces PPP, encapsulation Shim Layer.
   - Gestion des tunnels MPLS virtuels.
   - Espace d'étiquettes global ou spécifique à une interface.
   - Hiérarchie MPLS : prise en charge de la pile d'étiquettes.
   - Consultation d'étiquettes récursive *(ang. Recursive Label Lookups).*
   - Une entrée dans la table de routage du noyau Linux peut être associée à une étiquette de sortie. Ceci inclut le support pour un nombre variable de tables, TOS, et routage par interface.
   - Intégration avec le modèle de gestion de QoS Linux.
   - Support pour l'architecture DiffServ (E-LSPs, L-LSPs).
   - Support pour Ethernet sur MPLS.
2. *ldp-portable* : Une implémentation du protocole LDP.
   - Modes de distribution d'étiquettes non sollicité et sur demande *(ang. Unsollicited Downstream and Downstream on Demand).*
   - Gestion de Peers directes et indirectes.
   - Distribution d'étiquettes contrôlée par Politiques.
   - Intégration à la plateforme de routage Zebra.
   - API de configuration flexible (similaire à LDP-MIB).

Evolution Historique

Le projet MPLS pour Linux a débuté vers 1999, d'abord comme outil d'analyse pour le protocole LDP, suite à la publication par Nortel Networks d'une bibliothèque de fonctions pour la manipula-

tion, l'encodage et le décodage des PDUs. Le principal responsable du projet, James Leu, continue activement son développement. Depuis Avril 2000 environ, le logiciel est séparé en deux parties, comme indiqué ci-dessus.

Le projet MPLS pour Linux évolue rapidement. La liste de diffusion associée est très active (mpls-linux-general :sourceforge.net). Il reste encore un certain nombre de fonctionnalités à implémenter, notamment le protocole RSVP-TE, d'autant plus que le groupe de travail de l'IETF pense sérieusement à ne conserver que ce dernier.

Le lecteur est invité à consulter la page Web du projet pour plus d'informations.

Liste de publications

[1] R. Casellas, D. Kofman, *«MPLS, MultiProtocol Label Switching, from IP forwarding to label switching, support de cours »*, ENST 2000-2002INFRES74.

[2] R. Casellas, D. Kofman, J.L. Rougier, *«Load Sharing Schemes in MPLS networks»*, in 2nd European Conference in Universal Multiservice Networks, ECUMN'2002, Colmar, France.

[3] R. Casellas, *«Performance Evaluation of MPLS Load Sharing»*, in Workshop on Design and Performance Evaluation of 3G Internet Technologies, Fort Worth, Texas, 2002.

[4] R. Casellas, D. Kofman, *«Performance Evaluation of MPLS Load Sharing, Extended version»*, Suite à invitation du comité de programme de MASCOTS 2002, à soumettre pour publication dans édition spéciale de Performance Evaluation, 2002-2003.

[5] R. Casellas, D. Kofman, *«Adaptive Load Sharing»*, à soumettre.

Stages Encadrés

[1] Basheer Dargham, *«MPLS Optimized Multi-Path (MPLS-OMP), analyse et expérimentation»*, Stage ,2000.

[2] Halim Adel, *«Ingénierie de trafic dans les réseaux MPLS»*, Stage école polytechnique, 2001.

Bibliographie

[1] G. Apostolopoulos, R. Guérin, S. Kamat, A. Orda, and S. K. Tripathi, *«Intra-Domain QoS routing in IP Networks : A feasibility and Cost/Benefit Analysis»*, U.Maryland, U. Pennsylvania, Lucent Bell Labs, Technion I.I.T, U. CA, -.

[2] G. Apostolopoulos, R. Guérin, S. Kamat, and S. K. Tripathi, *«Quality of Service Based Routing : A Performance Perspective»*, U. Maryland, IBM T.J. Watson Res. Center, U. California, -.

[3] G. Apostopoulos and S. K. Tripathi, *«On Reducing the Processing Cost of On-demand QoS Path Computation»*, IEEE pags. 80-89, 1998. 4.6.3

[4] G. Apostolopoulos and S. K. Tripathi, *«On the Effectuveness of Path-Precomputation in Reducing the Processing Cost of On-Demand QoS Path Computation»*, IEEE, 1998.

[5] A. Orda, *«Routing with End-to-End QoS Guarantees in Broadband Networks»*, IEEE/ACM Transactions on Networking, Vol. 7, No. 3, June 1999. 4.5.1, 4.6.3

[6] B. Awerbuch, Y. Du, and Y. Shavitt, *«The Effect of Network hierarchy Structure on Performance of ATM PNNI Hierarchical routing»*, Dept. C.S. John Hopkins University, Bell Labs, Lucent Technologies, IEEE, 1998. 4.5.1

[7] S. Bahk and M. El Zarki, *«Dynamic Multi-path Routing and How it compares with other Dynamic Routing Algorithms for High Speed Wide Area Networks»*, COMM'92 ACM, 1992.

[8] J. E. Baker, *«A distributed Link restoration algorithm with Robust Preplanning»*, IEEE Globecom'91, 1991.

[9] C. S. Beightler, D. T. Phillips, and D. J. Wilde, *«Foundations of Optimization»*, Prentice-Hall International Series in Industrial ans Systems Engineering. ISBN 0-13-330332-2, 1979.

[10] D. Bertsekas and R. Gallager, *«Data Networks»*, Prentice Hall International Editions. ISBN 0-13-196981-1, 1987. 4.10.5, 6.4, 6.5

[11] D. D. Botvitch and N. Duffield, *«Large deviations, the shape of the loss curve, and economies of scale in large multiplexers»*, Queueing Systems, 1995. 5.8.3

[12] Jamoussi et al., *«Constraint-Based LSP Setup using LDP»*, draft-ietf-mpls-cr-ldp-03.txt, work in progress, Internet Draft, IETF, October 1999. 2.2.6, 4.2, 4.3

[13] C.S. Chang, *«Performance Guarantees in Communication Networks»*, Springer Verlag, TNCS, 2000. 1, 8.1.2, 1, A.4

[14] S. Chen and K. Nahrsted, *«An Overview of Quality of Service Routing for Next-Generation High Speed Networks : Problems and Solutions»*, IEEE Network, November/December 1998. 4.6.3

[15] S. Chen and K. Nahrstedt, *«Distributed QoS routing with Imprecise State Information»*, Dept. CS. University of Illinois ; IEEE pags. 614-621, 1998.

[16] S. Chen, K. Nahrstedt, and Y. Shavitt, *«A QoS Aware multicast Routing Protocol»*, Infocom, 2000.

[17] I. Cidon, R. Rom, and Y. Shavitt, *«Analysis of Multi-Path Routing»*, IEEE/ACM Transactions on Networking Vol. 7 No.6, December 1999.

[18] C. Courcoubetis, V.A. Siris, and G.D. Stamoulis, *«Application and Evaluation of Large Deviation Techniques for Traffic Engineering in Broadband Networks»*, ACM SIGMETRICS, 1998. 6.4.1, 6.6.3

[19] C. Courcoubetis, A. Dimakis, and G. D. Stamoulis, *«Traffic Equivalence and Substitution in a Multiplexer»*, Institute of Computer Science, Crete Univ., Infocom 1999. 5.8.3.2, 6.4.1, 6.6.1

[20] C. Courcoubetis, V.A. Siris, and G. Stamoulis, *«Application of the many sources asymptotic and effective bandwidths for traffic engineering»*, Telecommunications Systems 12, 167-191 , 1999.

[21] H. Cramér, *«Sur un nouveau théoréme-limite de la théorie des probabilités»* In Actualités Scientifiques et Industrielles, n 736 in Colloque cnsacré à la théorie des probabilités, pp. 5-23, Hermann, Paris 1938. 5.8.1, A.2

[22] H. De Neve and P. Van Mieghen, *«A multiple quality of service routing algorithm for PNNI»*, Alcatel Corporate Research., Proceedings 1998 IEEE ATM workshop, May 26-29.

[23] A. Dembo and O. Zeitouni, *«Large Deviations Techniques and Applications»*, Jones and Bartlett Pubislers, Inc. ISBN 0-86720-291-2, 1993. 5.8.1, 8.1.1, A.1, 1, A.5.1

[24] E. Dinan, D.O. Awduche, and B. Jabbari, *«Optimal Traffic Partitioning in MPLS Networks»*, George Mason University, UUNET (MCI Worldcom), 1999. 4.10.5

[25] N.G. Duffield and N/ O'Connell, *«Large deviations and overflow probabilities for the general single server queue, with applications»*, Math. Proc. Cam. Phil. Soc., 118, 363-374 , 1995. 6.6.1, A.6.2

[26] D. A. Dunn, W. D. Grover, and M. H. MacGregor, *«Comparision of k-shortest paths and maximum flow routing for Network facility Restoration»*, IEEE Journal on selected areas in Communications, Vol. 12, N. 1, January 1994. 4.10.4.1

[27] R.S. Ellis, *«Large Deviations for a general class of random vectors»*,Ann. Probab. Vol. 12, pp. 1-12, 1984. A.3.4

[28] R.S. Ellis, *«Entropy, Large deviations and Statistical Mechanics»*, New York, Springer-Verlag 1985.

[29] A.I. Elwalid and D. Mitra, *«Effective bandwidth of general Markovian traffic sources and admission control of high speed networks»*, IEEE/ACM Transactions on Networking, 1, 1993.

[30] D.Y. Eun and N. Shroff, *«Network Decomposition in the Many-Sources Regime»*, Advances in Applied Probability, submitted, 2002. 8.1, 9.8

[31] E. Felstaine and R. Cohen, *«On the distribution of Routing Computation in Hierarchical ATM Networks»*, IEEE/ACM Transactions on Networking., 1999. 4.5.1

[32] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford, *«NetScope : Traffic Engineering for IP Networks»*, AT&T Labs-Research, IEEE Network, March/April 2000.

[33] S. Floyd and V. Jacobson, *«Synchronization of periodic routing messages»*, IEEE/ACM Transactions on Networking, vol.2, pp 122-136, April 1994.

[34] L.R. Ford and D.R. Fulkerson, *«Flows in Networks»*, Princeton University Press, Princeton, New Jersey, 1962.

[35] B. Fortz and M. Thorup, *«Internet Traffic Engineering by Optimizing OSPF weights»*, AT&T Labs-Research. Infocom 2000, 2000.

[36] J. Gärtner, *«On large deviations from invariant measure»*, Theory Prob. Appl., Vol. 22, pp. 24-39, 1977 A.3.4

[37] A. Banerjee, J. Drake, J. Lang, B. Turner, K. Kompella, and Y. Rekhter, *«Generalized Multiprotocol Label Switching : An overview of Routing and Management Enhancements»*, IEEE Communications Magazine, January 2001. 3.6

[38] A. Banerjee, J. Drake, J. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter, *«Generalized Multiprotocol Label Switching : An overview of Signaling Enhancements and Recovery Techniques»*, IEEE Communications Magazine, July 2001. 3.6

[39] O. Gerstel, I. Cidon, and S. Zaks, *«The Layout of Virtual Paths in ATM Networks»*, IEEE/ACM Transactions on Networking vol. 4 No. 6, December 1996.

[40] R.J. Gibbens, *«Traffic Characterization and Effective bandwidths for broadband network traces»*, University of Cambridge , -.

[41] A. Gueroui, L. Mokdad, and J. Ben-Othman, *«A new Multiservices Rerouting Algorithm»*, IEEE, June 19, 2000.

[42] R. Guérin and A. Orda, *«QoS-based Routing in Networks with Inaccurate Information : Theory and Algorithms»*, IBM T.J Watson Res. Center, Technion I.I.T., Infocom97.

[43] F. Hao and E. W. Zegura, *«On Scalable QoS routing : Performance Evaluation of Topology Aggregation»*, Georgia Tech ; Atlanta, -. 4.5.1

[44] F. Hao and E. W. Zegura, *«Scalability Techniques in QoS Routing»*, Georgia Tech, Atlanta, -. 4.5.1

[45] F. Hao, E. W. Zegura, and S. Bhatt, *«Performance of the PNNI Protocol in Large Networks»*, Georgia Tech, Atlanta, Bellcore and DIMACS, -. 4.5.1

[46] M. Heusse and Y. Kermarrec, *«Adaptive Routing and Load Balancing of Ephemeral Connections»*, ENST de Bretagne, 2000. 4.10.5

[47] J.Y. Hui, *«Resource allocation for broadband networks»*, IEEE Journal Selected Areas Communication 6, 1598-1608, 1988. 5.5, 5.5.1, A.2

[48] P. Newman, G. Minshall, and T. Lyon, *«IP switching : ATM under IP»*, IEEE/ACM Transactions on Networking 6(2) :117-129, April, 1998. 2.1

[49] A. Iwata, R. Izmailov, and B. Sengupta, *«Alternative Routing methods for PNNI networks with Partially Disjoint Paths»*, C & C Media Res.Labs, NEC Corporation, 1999.

[50] J.J. Garcia-Luna-Aceves, *«Loop Free Routing Using Diffusing Computations»*, IEEE/ACM Transactions on Networking Vol. 1, No. 1, February 1993.

[51] J.J. Garcia-Luna-Aceves and S. Murthy, *«A Path-Finding Algorithm for Loop-Free Routing»*, IEEE/ACM Transactions on Networking Vol. 5, No.1, February 1997.

[52] J.L. Rougier, *«Routage Dynamique Unicast et Multicast», Thèse de doctorat*, ENST, 1999. 2.1.1, 4.5.1, 4.5.1

[53] F.P. Kelly, *«Notes on effective bandwidths»*, In «Stochastic Networks : Theory and Applications» (Editors F.P. Kelly, S. Zachary and I.B. Ziedins) Oxford University Press , 1996. 5.5, 5.5.1, 5.8, 5.8.3, 5.8.3, 5.8.3.1, 6.6.1, 6.6.3, 7.2.4, 9.6.1, 10.3, A.3.1

[54] J.F.C. Kingman, *«Inequalities in theory of queues»*, J. Roy. Stat. Soc., Series B, 32 pp. 102-110, 1970. A.6.2

[55] L. Kleinrock, *«Queueing Systems Theory»*, Wiley Interscience, 1975. 4.10.5

[56] D. Kofman and M. Gagnaire, *«Réseaux Haut Débit : Tome I Réseaux ATM et réseaux locaux», 2ed.*, Dunod, 1999. 2.1.1

[57] Andersson et al., *«LDP Specification»*, draft-ietf-mpls-ldp-08.txt, work in progress, Internet Draft, IETF, 26 Jun 2000.

[58] N. Likhanov and R. R. Mazumdar, *«Cell loss asymptotics in buffers fed with a large number of independent stationary sources»*, 0-7803-4386-7/98 IEEE, 1998. (document), 5.8.1, 5.8.3, 5.8.3.1, 5.8.3.1, 5.8.1, 5.8.2, 5.8.3.1, 6.4, 6.6, 6.6.3, 6.9, 7.2.1, 7.5.1, 9.4

[59] Z. Liu, P. Nain, and D. Towsley, *«Exponential bounds with applications to call admission»*, J.A.C.M., vol. 44, no. 3, May 1997, pp. 366-394. A.6.2

[60] D. H. Lorenz and A. Orda, *«QoS Routing with Uncertain Parameters»*, IEEE/ACM Transactions on Netwroking, Vol. 6, No. 6 , December 1998. 4.5.1

[61] R.M. Loynes, *«The stability of a queue with Non-Independent Inter-Arrival and Service Times»*, Proc. Cambridge Philos. Soc. 58, pp. 497-520, 1962.

[62] I. Widjaja and A. ElWalid, *«MPLS Adaptive Traffic Engineering»*, work in progress, Internet Draft, IETF, August 1998. 4.10.5

[63] E.C. Rosen, A. Viswanathan, and R. Callon, *«Multi Protocol Label Switching Architecture»*, RFC 3031 IETF, January 2001.

[64] J. Heinanen, *«»*, IETF, FIXME. 2.2.5

[65] C. Villamizar, *«MPLS Optimized Multipath (MPLS-OMP)»*, UUNET. IETF, February 25, 1999. Internet draft, work in progress.. 4.10.5

[66] , *«A framework for Multiprotocol Label Switching Architecture»*, work in progress, Internet Draft, IETF, 1999.

[67] Q. Ma and P. Steenkiste, *«On path Selection for Traffic with Bandwidth Guarantees»*, Comp. Science department, Carnegie Mellon University, -. 4.6.3

[68] Q. Ma and P. Steenkiste, *«Supporting Dynamic Inter-Class Resource Sharing : A multi-class QoS Routing Algorithm»*, Cisco Sytems, C.S. Dept, Carnegie Mellon University, 1998.

[69] L. Massoulié, *«Théorie des Grandes Déviations», notes de Cours*, DEA Probabilités Appliqués Paris VI., 2000. (document), A.3.4.2

[70] I. Matta and A.U. Shankar, *«Dynamic routing of real-time virtual circuits»*, in Proceedings of IEEE International Conference on Network Protocols, pp. 132-139, 1996.

[71] P. Van Mieghem, *«Topology Information Condensation in Hierarchical Networks»*, Delt University, Alcatel Corporate Research, -. 4.5.1

[72] P. Van Mieghem, *«Routing in a hierarchical tructure»*, Alcatel Corporate Research, -. 4.5.1

[73] I. Norros, *«A storage model with self-similar input»*, Queueing Systems, Vol. 16, pp. 387-396, 1994. 5.4.7, 5.6.4, 5.8.2, 6.6.1, 6.6.1

[74] J. Moy, *«Open Shortest Path First»*, IETF, RFC 2328. 4.5

[75] D. Katz, D. Yeung, and K. Kompella, *«Traffic Engineering Extensions to OSPF, work in progress»*, Work in Progress, February 2001. 3.6.1

[76] C. Villamizar, *«OSPF Optimized Multi Path (OSPF-OMP)»*, work in progress, Internet Draft, IETF, February 1999.

[77] C. H. Papadimitriu and K. Steiglitz, *«Combinatorial Optimization»*, Prentice-Hall, 1982.

[78] G.R. Walsh, *«Methods of Optimization»*, John Wiley and Sons. ISBN 0-471-91924-1., 1975.

[79] A. Orda and R. Rom, *«Shortest Path and Minimum Delay Algorithms in Networks with time dependent edge-length»*, Dept. Electrical Engineering, Technion I.T.T, August 1989.

[80] S. Oueslati-Boulahia, *«Qualité de service et routage et routage des flots élastiques dans un réseau multiservice»*, Thèse de doctorat, Novembre 2000. 4.6.3

[81] O. Ozturk, R.R. Mazumdar, and N. Likhanov, *«Many Sources Asymptotics for Networks with Small Buffers»*, Preprint, submitted to Queueing Systems, 2002. 8.1

[82] , *«ATM Forum PNNI Specifications»*, ATM Forum, 1996. 4.5.1

[83] V. Paxson and S. Floyd, *«Why we don't know how to simulate the Internet»*, In Proceedings of the Winter Simulation Conference , December 1997.

[84] G. Apostolopoulos, D. Williams, S. Kamat, R. Guerin, A. Orda, and T. Przygienda, *«QoS routing mechanisms and OSPF extensions»*, RFC 2676, January 1998.

[85] Z. Zhang, C. Sanchez, B. Salkewicz, and E.S. Crawley, *«Quality of service Extensions to OSPF or quality of service path first routing(QOSPF), work in progress»*, Work in Progress, September 1997.

[86] R. Guérin, D. Williams, and A. Orda, *«QoS Routing Mechanisms and SPF extensions»*, GLOBECOM, 1997.

[87] P. Aukia, M. Kodialam, P. V. N. Koppol, T. V. Lakshman, H. Sarin, and B. Suter, *«RATES : A Server for MPLS Traffic Engineering»*, Bell Laboratories, Lucent Laboratories. IEEE Network, March/April 2000. 4.9, 4.10.6

[88] D. O. Awduche, J. Malcom, J. Agogbua, M. O'Dell, and J. McManus, *«Requirements for Traffic Engineering over MPLS»*, IETF, 1999. 4.2, 4.3

[89] C. Hopps and D. Thaler, *«Multipath Issues in Unicast and Multicast Next-Hop Selection»*, IETF, November 2000. 4.10.1, 4.10.2

[90] C. Hopps, *«Analysis of an Equal-Cost Multi-Path Algorithm»*, IETF, November 2000. 4.10.1

[91] E. Rosen, A. Viswanathan, and R. Callon, *«Multiprotocol Label Switching Architecture»*, IETF, Jan 2001. 4.10.7

[92] F. Le Faucheur, L. Wu, B. Davi, S. Davar, P. Vaanane, R. Krishna, P. Cheva, and J. Heinanen, *«Multi-Protocol Label Switching (MPLS) Support of Differentiated Services»*, IETF, Mai 2002. 4.10.7

[93] J. Postel, *«Internet Protocol»*, September 1981, IETF.

[94] J. Postel, *«Transmission Control Protocol»*, September 1981, IETF.

[95] D. O. Awduche et al., *«Extensions to RSVP for LSP Tunnels»*, draft-ietf-mpls-rsvp-lsp-tunnel-05.txt, work in progress, Internet Draft, IETF, February 2000. 2.2.6, 4.2, 4.3

[96] P. Rabinovitch, *«Statistical estimation of effective bandwidth»*, M.Sc. thesis, University of Cambridge, 2000. 5.4.7, 5.7, 5.7, 5.8.2

[97] D. S. Reeves and H. F. Salama, *«A Distributed Algorithm for Delay-Constrained Unicat Routing»*, IEEE/ACM Transactions on Networking Vol. 8, No. 2, April 2000.

[98] A. Shaik, J. Rexford, and K. G. Shin, *«Efficient precomputation of quality-of-service routes»*, in Proceedings of Workshop on Network and Operating System Support for Digital Audio and Video, July 1998. 4.6.3

[99] A. Shaikh, J. Rexford, and K. G. Shin, *«Load-Sensitive Routing of Long-Lived IP flows»*, Dept. of E.E. and C.S., university of Michigan. Network Mathematics Research, AT&T Labs, -.

[100] A. Shaikh, J. Rexford, and K. G. Shin, *«Evaluating the Impact of Stale Link-State on Quality-of-Service Routing»*, IBM T.J. Watson Res. Center ; AT&T Labs ; EECS Dept, U. Michigan, 1998.

[101] A. Simonian and J. Guibert, *«Large deviations approximations for fluid queues fed by a large number of on/off sources»*, IEEE JSAC, 13(7) :1017-1027 , August 1995. 5.8.3, 6.4

[102] W. K. Tsai, J. K. Antonio, and G. M. Huang, *«Complexity of Gradient Projection Method for Optimal Routing in Data Networks»*, IEEE/ACM Transactions on Networking, Vol.7 no. 6, December 1999. 4.10.5

[103] S. Vutukury and J.J. Garcia-Luna-Aceves, *«A simple Approximation to Minimum-Delay Routing»*, Baskin School of Engineering. University of California, 1999.

[104] Z. Wang and J. Crowcroft, *«Quality-of-Service Routing for supporting Multimedia Applications»*, IEEE Journal on selected areas in communications, vol.14 No.7, September 1996. 4.6.3

[105] C. Courcoubetis and R. Weber, *«Buffer overflow asymptotics for a switch handling many traffic sources»*, Journal of Applied Probability, 1996. 5.8.3, 6.6

[106] D. Wischik, *«The output of a switch, or, effective bandwidths for networks»*, Queueing Systems 32 :386-393, 1999. 8.1, 9.1

[107] D. Wischik, *«Sample path large deviations for queues with many inputs»*, Annals of Applied Probability, 2000.

[108] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee, *«How to model an internetwork»*, In Proceedings of IEEE INFOCOM,pp. 594-602 , March 1996.