INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Partage de bande passante dans un réseau : approches probabilistes

Guy Fayolle — Jean-Marc Lasgouttes (éditeurs)

N° 4202

Juin 2001

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4202--FR

Partage de bande passante dans un réseau : approches probabilistes

Guy Fayolle, Jean-Marc Lasgouttes (éditeurs)

Thème 1 — Réseaux et systèmes
Projet Meval

Rapport de recherche n° 4202 — Juin 2001 — 72 pages

Résumé : Ce rapport reprend une série d'études menées au sein du projet MEVAL– dans le cadre d'une consultation thématique de France Télécom R&D– sur les effets des politiques de partage de bande passante sur le comportement dynamique d'un réseau.

On s'intéresse à un réseau de type « best effort » où la bande passante disponible est distribuée de la manière équitable entre les utilisateurs. En particulier, on considère la politique dite « min », qui est une approximation pessimiste de la politique classique d'« équité max-min ». Dans le premier chapitre, on donne les conditions nécessaires et suffisantes d'ergodicité, ainsi que divers résultats de grandes déviations.

Le second chapitre est consacré à une analyse poussée du système en limite thermodynamique, en particulier à partir d'approximation champ moyen. Une nouvelle approche de la propagation du chaos est proposée pour les systèmes non symétriques

Mots-clés : réseaux « best-effort », équité max-min, politique min, fonctions de Lyapounov, grandes déviations, fonctions analytiques, champ moyen

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Bandwidth sharing in a network: probabilistic studies

Abstract: This report summarizes a series of studies—partly supported by a grant from France Télécom R&D—carried out in the Meval project, and which concern bandwidth sharing policies in data networks.

One considers more specifically the so-called *min* policy, which is a conservative approximation to the classical *max-min* policy. In the first chapter, necessary and sufficient conditions are established for a wide class of best-effort networks, together with large deviations results.

The second chapter presents a rather extensive analysis of the models in thermodynamic limit. Some new approach to propagation of chaos is proposed for non symmetrical systems.

Key-words: best-effort networks, max-min fairness, min policy, Lyapounov functions, large deviations, analytic functions, mean field

Avant propos

Le présent rapport est un document de synthèse contenant l'essentiel des travaux réalisés au titre d'une action contractuelle passée entre France Télécom R&D et l'INRIA. Il s'agissait principalement d'étudier les problèmes d'ingénierie du trafic, liés à l'introduction du service ABR, afin de dégager des règles de dimensionnement et de contrôle de flux pour les réseaux offrant un tel service.

Le document comprend deux volets principaux :

- le premier (chapitre 1) présente des résultats probabilistes pour des topologies de réseau assez générales ;
- le deuxième (chapitre 2) est consacré aux études en *limite thermodynamique*, c'est-à-dire lorsque la taille des systèmes augmente.

Une partie de ces recherches (en particulier les sections 1.2, 1.3, 2.1 et 2.2) a fait l'objet d'une publication orientée « télécom » [6]. La section 2.3 est la substance de l'article [8], au contenu plus théorique.

Les travaux ont été réalisés au sein du projet MEVAL de l'INRIA, à l'aide de méthodes de modélisation mathématique, les résultats obtenus étant le fruit du travail conjoint de chercheurs de divers organismes :

- Franck Delcoigne, Guy Fayolle et Jean-Marc Lasgouttes (INRIA, projet MEVAL) ;
- Arnaud de La Fortelle (ENPC, mis à disposition à l'INRIA par le ministère de l'équipement) ;
- Thomas Deneux (stagiaire de l'ENS Rue d'Ulm).

Au début de chaque section, sont mentionnés les auteurs plus spécifiques à celle-ci.

Table des matières

1 Résultats probabilistes généraux

On s'intéresse dans ce chapitre à un réseau gérant des flots de données pour plusieurs utilisateurs, dans lequel aucune garantie de service n'est offerte. Le modèle de service le plus raisonnable dans ce contexte est le service dit « best effort », qui alloue la bande passante de manière équitable entre les utilisateurs actifs (voir par exemple Bertsekas et Gallager [1]). En fait, il y a plusieurs notions d'équité dans ce genre de problème, la plus classique étant l'équité max-min.

Même si on commence à bien comprendre la manière dont la bande passante est répartie par les différents mécanismes de contrôle de congestion, assez peu de travaux permettent de rendre compte de l'impact de la nature aléatoire du trafic, notamment sur la qualité de service perçue. À notre connaissance, une des seules étude en ce sens est celle de Roberts et Massoulié [12], qui analysent une topologie de réseau linéaire.

Notre propos est d'étudier des topologies plus complexes sous différents angles. Un modèle markovien assez général de réseau « best effort » est présenté en section 1.1 ; il est destiné à rendre compte de l'impact de la topologie, des paramètres de trafic et de la politique de partage de bande passante sur la performance. Des résultats de simulation d'une version du réseau en étoile sont donnés en section 1.2. Les conditions d'ergodicité de ce réseau sont considérées dans la section 1.3. Enfin, la section 1.4, à l'aide de techniques de *grandes déviations*, donne des idées sur la façon *optimale* de saturer le réseau et met en évidence les possibles goulots d'étranglement.

1.1 Présentation du modèle

On considère un réseau comprenant un ensemble $\mathcal{L}$ de liens. Plusieurs sources établissent des connexions le long de routes qui empruntent ces liens. On s'intéresse à la façon de partager la bande passante offerte par chaque lien entre les connexions actives, ainsi qu'à l'effet de la politique d'allocation sur la dynamique du réseau. Un exemple simple d'un tel modèle est un réseau en étoile (figure 1.1.1) où toutes les routes sont de longueur 2, qui représente assez bien un gros commutateur avec de nombreuses routes. Chaque branche de l'étoile comporte alors deux liens (« entrant » et « sortant ») et chaque route est isomorphe à un couple de branches.

Soit $\mathcal{R}$ l'ensemble des routes du réseau. On note alors L_r la longueur d'une route $r \in \mathcal{R}$ et $r(1), \dots, r(L_r)$ les liens qu'elle emprunte. Les connexions sont créées sur cette route selon un processus de Poisson d'intensité λ_r. Chaque connexion est maintenue jusqu'à avoir transmis à travers le réseau des données dont le volume suit une loi exponentielle de moyenne σ_r. Un lien

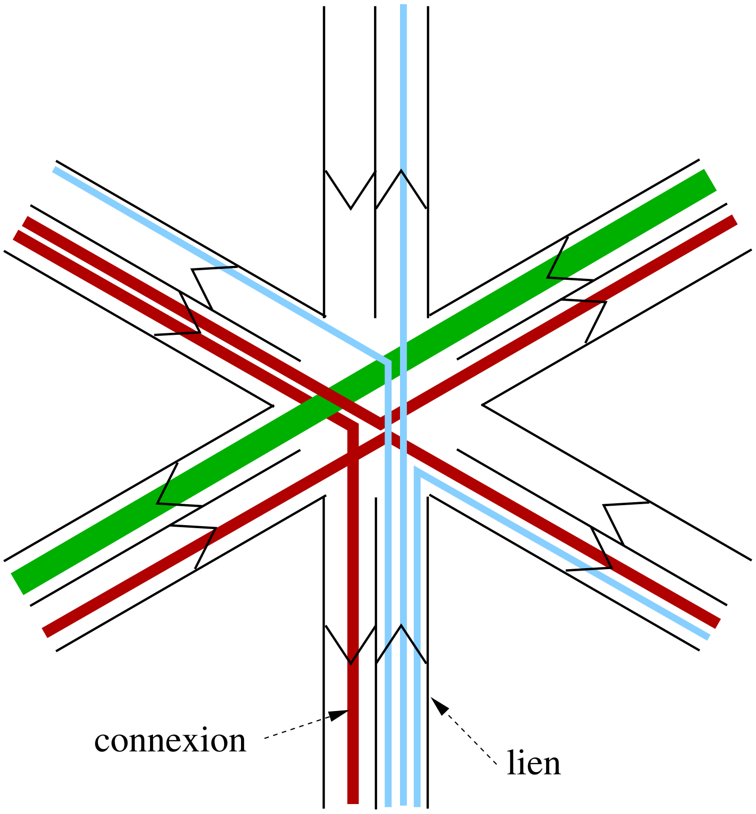

FIG. 1.1.1: *Réseau en étoile.*

$\ell \in \mathcal{L}$ a une bande passante C_ℓ et l'intensité d'arrivée à ce lien est notée

$$\rho_\ell \stackrel{\text{def}}{=} \frac{1}{C_\ell} \sum_{r \ni \ell} \lambda_r \sigma_r,$$

la sommation étant effectuée sur toutes les routes utilisant le lien ℓ. Chaque route r dispose à chaque instant d'une fraction ζ_r de la bande passante disponible.

L'état du système à un instant $t \in \mathbb{R}$ est défini par le nombre de connexions actives sur chaque route $\big(x_r(t),\, r \in \mathcal{R}\big) \stackrel{\text{def}}{=} X(t)$, et la quantité

$$X_\ell(t) \stackrel{\text{def}}{=} \sum_{r \ni \ell} x_r(t)$$

représente le nombre de connexions sur le lien $\ell \in \mathcal{L}$.

Le vecteur $X(t)$ est un processus de Markov dont les transitions possibles sont :

- arrivée d'une connexion sur une route r, avec intensité λ_r ;
- départ d'une connexion sur une route r, avec intensité $\zeta_r(X(t))/\sigma_r$.

Le générateur infinitésimal du processus s'écrit, pour une fonction test $f : \mathbb{N}^{\mathcal{R}} \mapsto \mathbb{R}$,

$$\begin{aligned} Gf(X) \stackrel{\text{def}}{=} & \sum_{r \in \mathcal{R}} \lambda_r \big[f(X + e_r) - f(X)\big] \\ & + \sum_{r \in \mathcal{R}} \frac{x_r}{\sigma_r} \zeta_r(X) \big[f(X - e_r) - f(X)\big], \end{aligned}$$

où e_r est le r-ième vecteur de base de $\mathbb{R}^{\mathcal{R}}$.

Jusqu'ici, nous n'avons pas discuté la manière de répartir la bande passante disponible entre les connexions en présence. Si $\zeta = (\zeta_r, r \in \mathcal{R})$ désigne une allocation quelconque réalisable, il vient alors nécessairement

$$\sum_{r \ni \ell} x_r(t)\zeta_r(X(t)) \leq C_\ell, \quad \forall \ell \in \mathcal{L}, \tag{1.1.1}$$

avec, par convention, $\zeta_r(X(t)) = 0$ lorsque $x_r(t) = 0$.

Un candidat assez naturel est le principe d'*équité max-min* (voir par exemple Bertsekas et Gallager [1]) : l'allocation ζ^{eq} est telle qu'il existe toujours un lien ℓ_r tel que

$$\sum_{r' \ni \ell_r} c_{r'}(t)\zeta^{\text{eq}}_{r'}(X(t)) = C_{\ell_r}, \quad \zeta^{\text{eq}}_r(X(t)) = \max_{r' \ni \ell_r} \zeta^{\text{eq}}_{r'}(X(t)). \tag{1.1.2}$$

Même si cette condition assure l'unicité de ζ^{eq} pour tout système fini, elle mène à des calculs relativement inextricables, dépendant de manière globale de l'état du réseau. Nous étudierons donc principalement la politique dite du « min », proposée par L. Massoulié et J. Roberts [11], où une connexion sur r se voit attribuer un débit

$$\zeta^{\min}_r(X(t)) \stackrel{\text{def}}{=} \min_{\ell \in r} \frac{C_\ell}{X_\ell(t)}. \tag{1.1.3}$$

Il est facile de voir que $\zeta^{\min}$ satisfait (1.1.1) et que cette politique « min » est toujours sous-optimale par rapport à l'équité max-min, au sens du lemme ci-après

Lemme 1.1.1. *À conditions initiales égales, le vecteur $X^{\text{eq}}(t)$ du système évoluant sous le principe d'allocation ζ^{eq} est stochastiquement inférieur au vecteur $X^{\min}(t)$ correspondant à $\zeta^{\min}$.*

Preuve. Supposons que, pour un t donné, on ait $X^{\text{eq}}(t) \leq X^{\min}(t)$. Alors, avec les notations de l'équation (1.1.2),

$$\begin{aligned}
\zeta^{\text{eq}}_r(X^{\text{eq}}(t)) &= \max_{r' \ni i_r} \zeta^{\text{eq}}_{r'}(X^{\text{eq}}(t)) \\
&\geq \frac{1}{X^{\text{eq}}_{i_r}(t)} \sum_{r' \ni i} x^{\text{eq}}_{r'}(t)\zeta^{\text{eq}}_{r'}(X^{\text{eq}}(t)) \\
&= \frac{C_{i_r}}{X^{\text{eq}}_{i_r}(t)} \\
&\geq \frac{C_{i_r}}{X^{\min}_{i_r}(t)} \\
&\geq \zeta^{\min}_r(X^{\min}(t)).
\end{aligned}$$

Il est donc facile de construire par couplage les processus $(X^{\text{eq}}(t), t > 0)$ et $(X^{\min}(t), t > 0)$ de façon à assurer que $X^{\text{eq}}(t) \leq X^{\min}(t)$ pour tout $t > 0$. ∎

Une dernière propriété agréable (mais qui ne sera pas directement utilisée par la suite) de la politique min est qu'elle est monotone par rapport à ses paramètres, au sens du lemme suivant.

Lemme 1.1.2. *Au sens de la monotonie stochastique, pour tout $t > 0$, $X^{\min}(t)$ est*

- *croissant avec la condition initiale $X^{\min}(0)$,*
- *croissant avec les taux d'arrivée λ_r, $r \in \mathcal{R}$,*
- *croissant avec les volumes moyens de données σ_r, $r \in \mathcal{R}$,*
- *décroissant avec les capacité de liens C_ℓ, $\ell \in \mathcal{L}$.*

Preuve. On utilise le fait que $\zeta_r^{\min}$ est une fonction décroissante de X. Cela permet, comme dans le lemme précédent, de construire des couplages majorants (ou minorants). Ce raisonnement ne serait pas possible pour la politique max-min, puisque ζ^{eq} n'a pas les propriétés requises (voir par exemple [1]). ■

Enfin, la section 2.6.2 étudie une classe de politiques dites « efficaces », qui contient notamment l'équité max-min et la politique d'équité proportionnelle introduite par Kelly [10] :

Définition 1.1.3. *Une politique est dite* efficace *si elle possède les propriétés suivantes :*

1. *pour toute route r, il existe $\ell_r \in r$ tel que*

$$\sum_{r' \ni \ell_r}^{n} x_{r'}(t)\tilde{\zeta}_{r'}(X(t)) = C_{\ell_r} \; ; \tag{1.1.4}$$

2. *pour un lien ℓ fixé, si l'allocation $\tilde{\zeta}$ obtenue en supprimant la contrainte de capacité au lien ℓ vérifie*

$$\sum_{r \ni \ell}^{n} x_r(t)\tilde{\zeta}_r(X(t)) \leq C_\ell, \tag{1.1.5}$$

   alors $\zeta_r = \tilde{\zeta}_r$ pour tout $r \ni \ell$.

Cette classe de politiques comprend celles qui optimisent une fonctionnelle sous contrainte de capacité (équité proportionnelle, ...), mais aussi l'équité max-min.

1.2 Simulation de réseaux en étoile[1]

On considère un réseau en étoile dans lequel la bande passante est partagée entre les liens selon une politique min ou max-min. La complexité mathématique de ces modèles impose de les simuler pour faire apparaître leurs propriétés qualitatives, dont les résultats corroborent certaines hypothèses (par exemple la propagation du chaos de la section 2.1).

1. Auteur : A. de La Fortelle

Nous avons procédé à des séries de simulations sur des séries de réseaux symétriques avec un nombre de liens variable (3 à 300), pour différentes valeurs de la charge ρ.

Les notations suivantes seront utilisées dans la description des algorithmes :

- $d_{\ell\ell'}^{(N)}$: allocation de débit pour une connexion (ℓ, ℓ) ;
- $d_\ell^{(N)} \stackrel{\text{def}}{=} \sum_{\ell'} c_{\ell\ell'} d_{\ell\ell'}^{(N)}$: débit d'un lien ℓ ;
- $D^{(N)} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\ell\ell'} c_{\ell\ell'} d_{\ell\ell'}^{(N)}$: débit moyen.

1.2.1 Simulation de la politique min

1.2.1.1 Algorithme

L'algorithme de la simulation est le suivant :

Initialisation Chaque lien ℓ reçoit $\tilde{x}_\ell$ connexions, les $\tilde{x}_\ell$ sont des variables aléatoires indépendantes de loi $X/2$, où X représente le champ moyen, dont une solution numérique a été calculée[2]. Ces $\tilde{x}_\ell$ connexions sont réparties uniformément, pour chaque lien ℓ, sur les $N-1$ autres liens.

Boucle À chaque itération on procède au tirage d'un saut du processus de Markov sous-jacent.

Calcul du temps de saut Le temps de saut est une variable exponentielle de paramètre $\lambda + D^{(N)}$,

Calcul des statistiques On enregistre les variables représentatives du réseau, essentiellement la distribution du nombre de connexions par lien.

Tirage du saut Le tirage du saut correspond à une implémentation particulière du générateur du processus de Markov. Une nouvelle connexion (ℓ, ℓ') apparaît avec intensité $\lambda/(N-1)$ et une connexion (ℓ, ℓ') est servie avec intensité $1/(X_\ell^{(N)} \vee X_{\ell'}^{(N)})$, où $X_\ell^{(N)}$ est le nombre de connexions sur le lien ℓ. On tire une arrivée (resp. un départ) avec probabilité $\lambda/(\lambda + D^{(N)})$ (resp. $D^{(N)}/(\lambda + D^{(N)})$).

- *Arrivée* : la connexion d'arrivée est tirée uniformément sur les paires (ℓ, ℓ') avec $\ell \neq \ell'$;
- *Départ* : on tire un lien ℓ avec une probabilité proportionnelle au débit $d_\ell^{(N)}$ du lien ℓ ; puis, on tire une connexion (ℓ, ℓ') avec une probabilité proportionnelle à $x_{\ell\ell'}^{(N)}/(X_\ell^{(N)} \vee X_{\ell'}^{(N)})$ (débit des $x_{\ell\ell'}^{(N)}$ connexions entre ℓ et ℓ').

Remise à jour des débits Les débits $d_\ell^{(N)}$ et $D^{(N)}$ sont nécessaires pour le tirage des sauts : il faut donc les remettre à jour.

2. Ces valeurs initiales sont utiles pour accélérer la convergence vers l'état stationnaire, mais ne changent pas les résultats. Le calcul numérique du champ moyen est présenté en section 2.2

Résultats Trois types de résultats sont transcrits sur fichiers : d'une part une partie des trajectoires du systèmes (les $X_\ell^{(N)}(t)$), et d'autre part les statistiques recueillies, c'est-à-dire la distribution des $X_\ell^{(N)}$ et la distribution du total $\sum_\ell X_\ell^{(N)}$.

1.2.1.2 Résultats de la simulation

Les résultats des simulations sont présentés dans les figures 1.2.1–1.2.3.

On doit remarquer que les valeurs données aux liens lors de l'initialisation ne conduisent évidemment pas à considérer que le réseau est en régime stationnaire. D'abord, lorsque N est petit, la distribution d'équilibre est loin du champ moyen ; ensuite, les liens, même si ils tendent à devenir indépendants deux à deux, ne sont absolument pas mutuellement indépendants.

Les calculs sont entachés d'une certaine imprécision, car l'expérience montre que la convergence vers l'état stationnaire est très lente, beaucoup plus lente que l'aspect « lisse » des distributions $X_\ell^{(N)}$ ne peut le faire croire.

1.2.1.3 Discussion des résultats

Il est bien sûr toujours risqué de faire des déductions à partir de graphiques, et les courbes obtenues seront surtout utilisées comme guide pour la suite de cette étude. En particulier, on peut remarquer les point suivants :

– La mesure stationnaire du nombre de connexions par lien croît avec le nombre de liens N (voir figure 1.2.2). Ceci suggère qu'il est possible d'étudier les suites de réseaux par monotonie, et permet aussi d'avoir des bornes supérieures sur la charge du réseau. De plus, on verra en section 2.2 que le modèle en champ moyen est très proche du modèle limite quand N est grand (ce qui justifie a posteriori les heuristiques qui ont été utilisées) et fournit une borne supérieure facile à évaluer.

– Lorsqu'il s'approche de la saturation ($\rho \to 1$), le réseau a un fonctionnement modal très marqué. Les arguments théoriques développés plus loin en section 2.3 laissent prévoir que ce mode serait en $C/(1-\rho)^2$, et on mesure sur les résolutions numériques que la moyenne est en $0{,}75/(1-\rho)^2$. On met ainsi le doigt sur un comportement qui est fondamentalement différent des files d'attente (M/M/1,...) classiques et mêmes des réseaux de files d'attente (Jackson,...).

1.2.2 Simulation de la politique max-min

1.2.2.1 Algorithme

L'algorithme de la simulation est semblable à celui du min (cf. section 1.2.1.1), mais la nature de la politique le rend plus complexe et surtout beaucoup plus lent. Les différences sont, essentiellement :

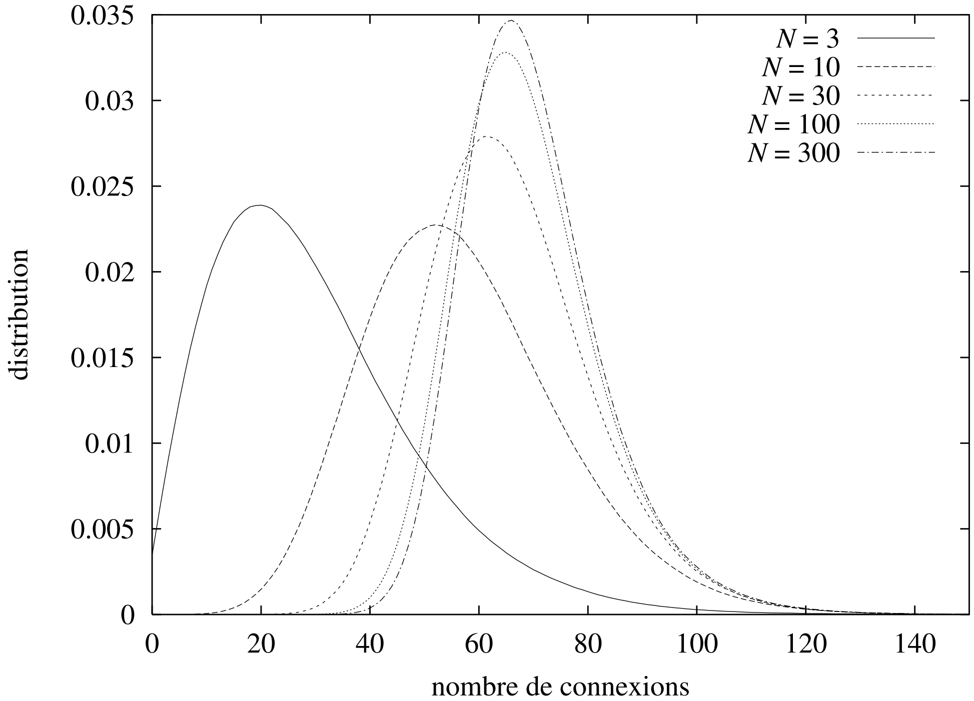

FIG. 1.2.1: *min : simulation d'un réseau en étoile pour $\rho = 0{,}90$ et N variable. La convergence vers un modèle limite est très prononcée.*

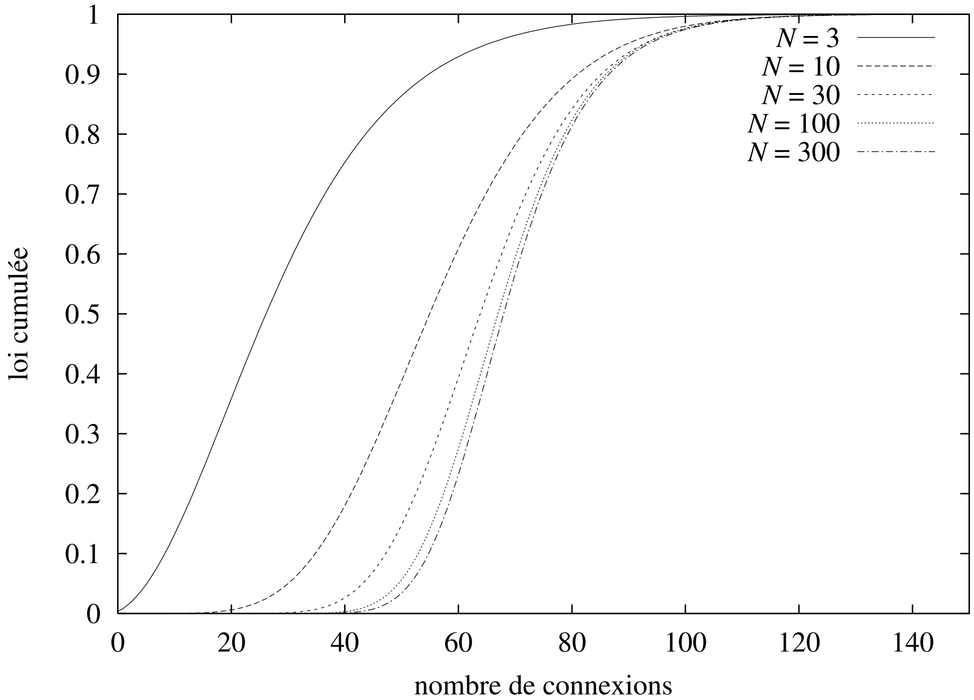

FIG. 1.2.2: *min : simulation d'un réseau en étoile pour $\rho = 0{,}90$ et N variable (lois cumulées). Les lois semblent croissantes avec N. (10^8 itérations par simulation)*

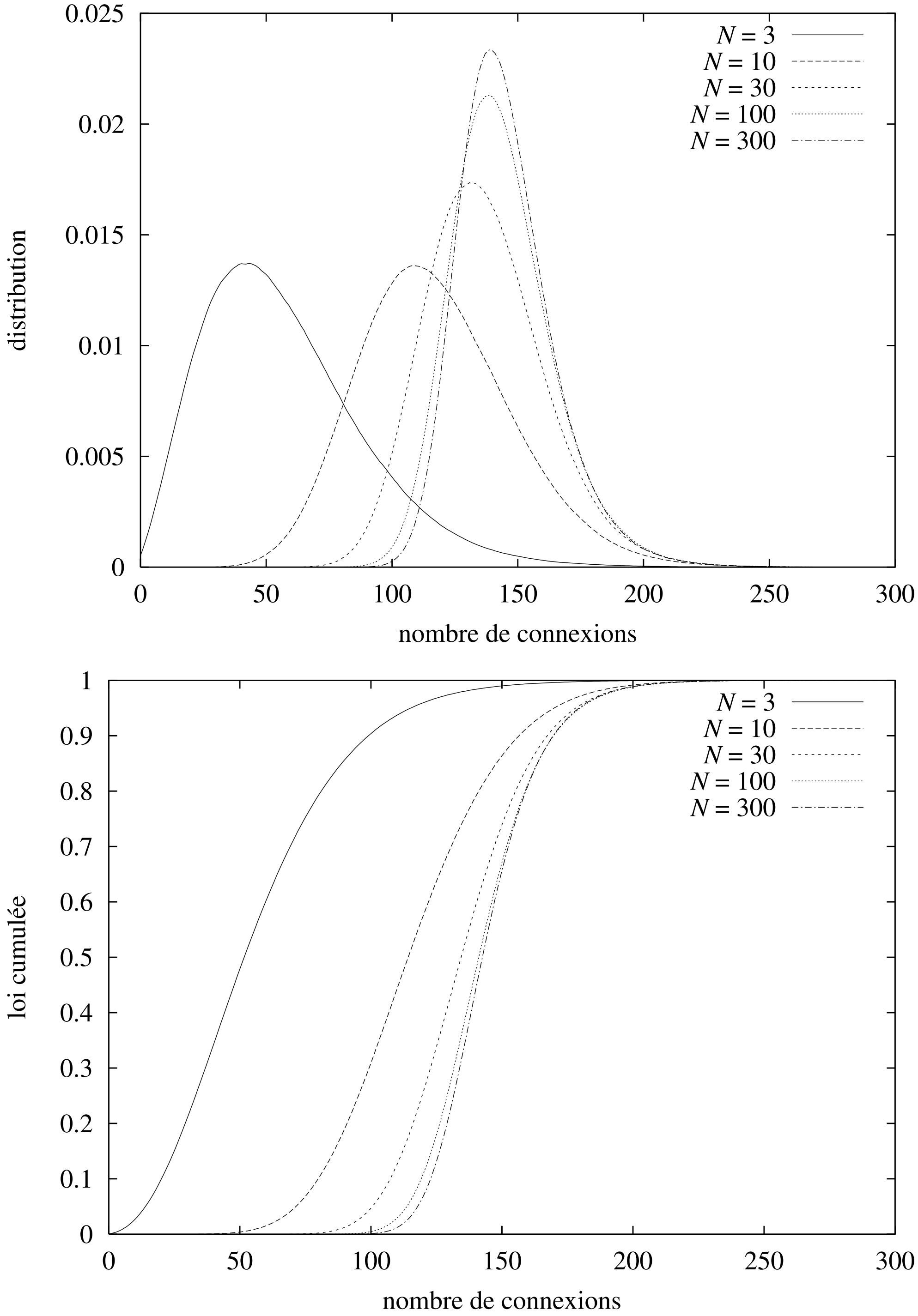

FIG. 1.2.3: *min : simulation d'un réseau en étoile pour $\rho = 0{,}93$ et N variable. Pour cette charge, le nombre moyen (en champ moyen) de connexions par lien est* 147,8. *On constate une convergence croissante assez bonne, même si, avec* 300 *liens, la corrélation entre liens est encore de* 0,0722 *(et donc non négligeable). Chaque simulation comprend* 10^8 *tirages.*

Initialisation Nous ne savons pas calculer de politique en champ moyen ; le réseau est donc initialisé à zéro.

Boucle Comme on ne sait pas recalculer rapidement les débits $d_{\ell\ell'}^{(N)}$, pour ℓ donné, il faut donc stocker et recalculer *tous* les débits $d_{\ell\ell'}^{(N)}$ à chaque remise à jour.

En notant d_ℓ^r [resp. X_ℓ^r] le débit [resp. connexions] restant (initialisé à 1 [resp. X_ℓ]), l'algorithme de calcul des débits exécute la boucle suivante :

- calcul du débit minimal par connexion $d_\ell = d_\ell^r / X_\ell^r$,
- calcul du lien limitant ℓ tel que $d_\ell = \min_k d_k$,
- pour tous les liens ℓ' restants passant par ℓ :
  - allocation du débit $d_{\ell\ell'} = d_{\ell'\ell} = d_\ell$,
  - remise à jour des débits restants $d_{\ell'}^r \leftarrow d_{\ell'}^r - x_{\ell\ell'} d_\ell$,
  - remise à jour des connexions restantes $X_{\ell'}^r \leftarrow X_{\ell'}^r - x_{\ell\ell'}$.
- élimination du lien ℓ ; on continue tant qu'il reste des connexions ($X_\ell^r > 0$).

La différence majeure par rapport à l'algorithme min réside dans la complexité du calcul des débits $d_{\ell\ell'}^{(N)}$, qui conduit à une complexité en $O(N^2)$ contre $O(N)$.

Au cours du développement, les algorithmes ont été soigneusement optimisés, ce qui a permis de passer (pour min), de simulations avec 10^7 tirages à 10^9 (ce qui représente quelques heures sur une station de travail), alors que dans le cas max-min, 98% du temps est passé à recalculer des débits, en dépit des améliorations, et on doit se limiter à 10^7 tirages.

1.2.2.2 Résultats

Les résultats des simulations sont présentés dans les figures 1.2.4 et 1.2.5. On doit remarquer que l'initialisation à zéro ne fausse pas les statistiques, puisqu'on ne considère que les 90 derniers pourcents de la simulation, et que le système est alors chargé.

1.2.2.3 Discussion des résultats

Avec les mêmes réserves que dans le cas de la politique min sur les dangers de la déduction fondée sur des graphiques, la simulation semble montrer les points suivants

- Un mode apparaît également dans la politique max-min, mais de façon nettement moins prononcée, puisque la moyenne (le lieu du pic), est proportionnel à $1/(1-\rho)$ lorsque le système se charge ($\rho \to 1$), contre $1/(1-\rho)^2$ pour le min. On note que la moyenne d'une file M/M/1 est justement $1/(1-\rho)$, et que la politique max-min n'en diffère donc que par une constante. Ainsi, même si la politique max-min n'est pas optimale, elle n'en est pas aussi éloignée que la politique min.

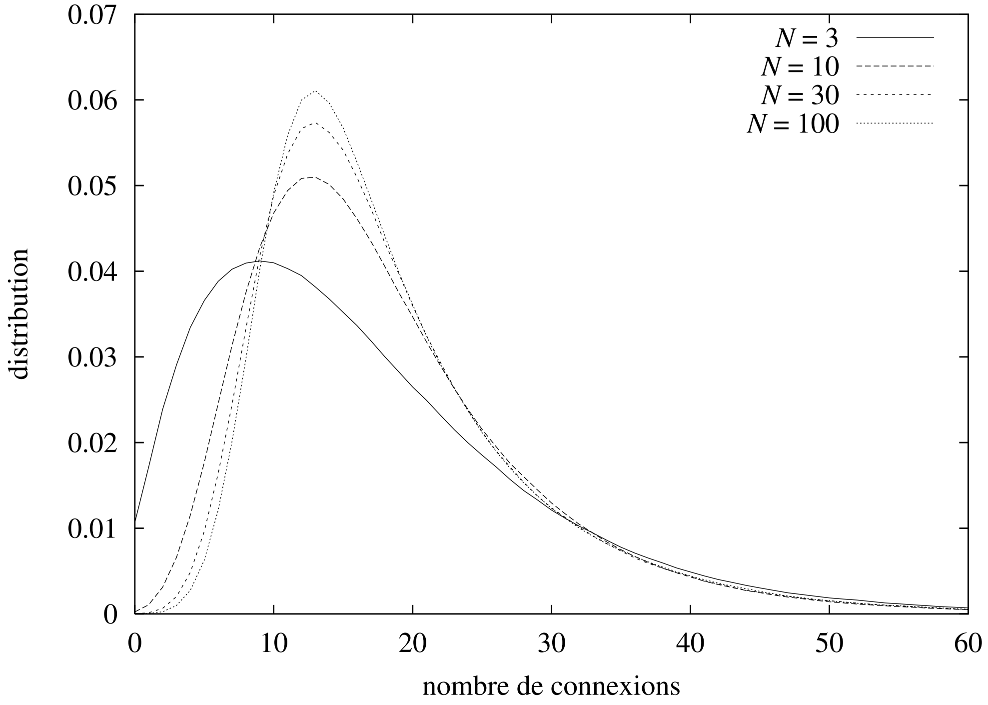

FIG. 1.2.4: *max-min : simulation d'un réseau en étoile pour $\rho = 0{,}90$ et N variable. La convergence vers un champ moyen est prononcée.*

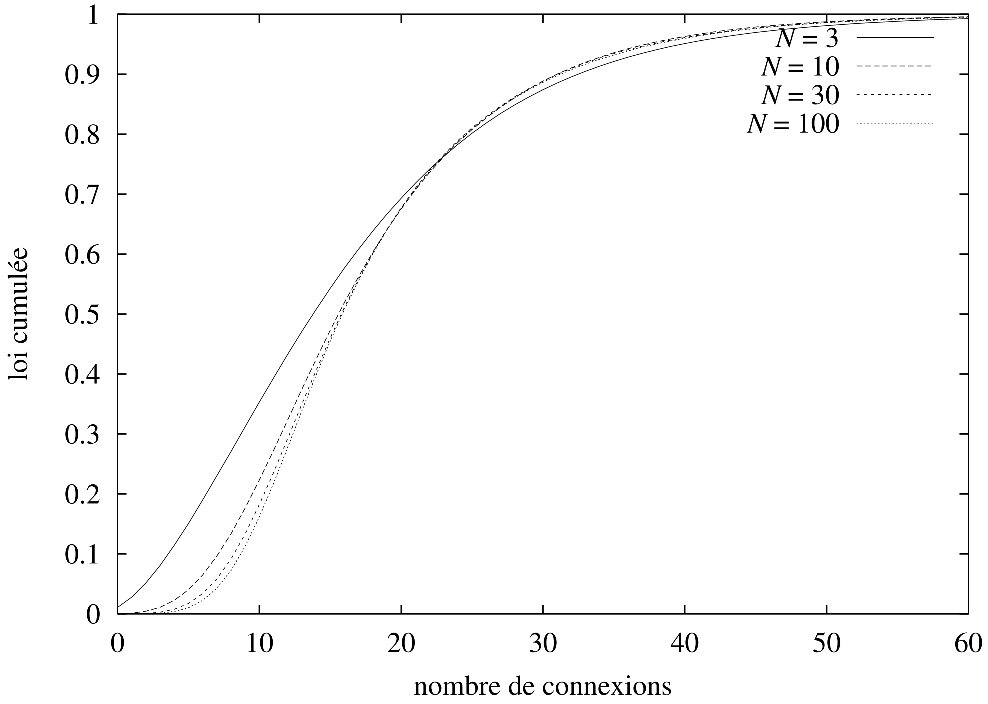

FIG. 1.2.5: *max-min : simulation pour $\rho = 0{,}90$ et N variable (lois cumulées). La croissance en N ne semble pas aussi évidente que dans le cas du min (10^7 itérations par simulation)*

– Bien que le nombre moyen de connexions semble croître avec le nombre de liens, cette croissance n'apparaît pas sur les lois cumulées (figure 1.2.5). Comme il n'y a pas non plus de monotonie stochastique sous la politique max-min, la convergence vers une loi de champ moyen risque d'être très difficile à démontrer.

– Les calculs sont entachés d'une certaine imprécision, car l'expérience montre que la convergence vers l'état stationnaire est très lente. Bien que les simulations comprennent environs 10 millions de sauts, la précision des distributions obtenues n'est que de l'ordre de 1%. Cette constatation est largement corroborée par la donnée de la statistique du nombre total de connexions, dont on peut voir sur la figure 1.2.6 qu'elle ne converge que très lentement vers la gaussienne de mêmes moyenne et variance. Il a fallu 14h (sur une station Compaq PWS500 à 500 MHz) pour réaliser cette simulation.

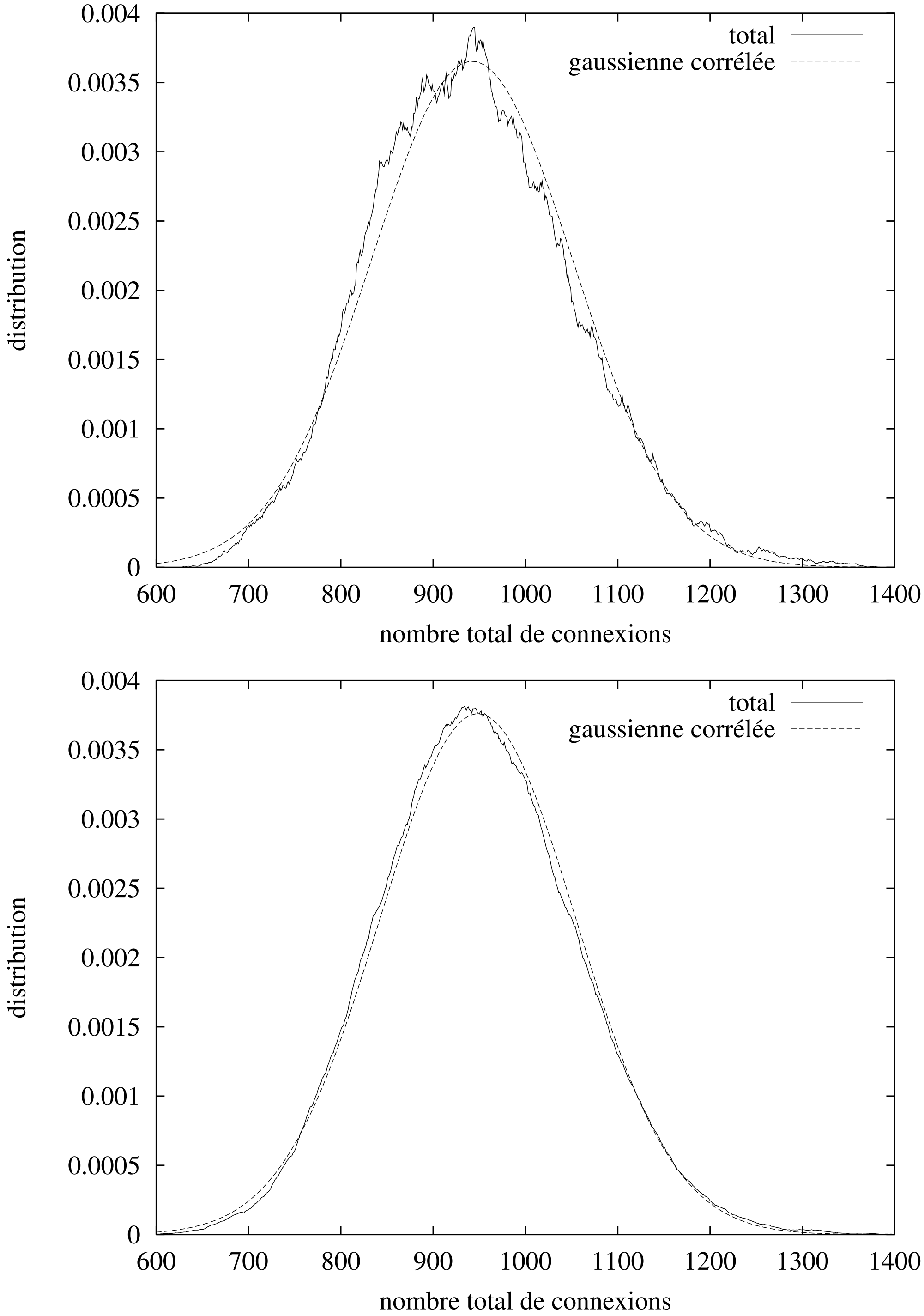

FIG. 1.2.6: *max-min: distribution du total de connexions, après* 10^7 *itérations (resp.* 10^8*, en bas), pour une étoile à cent liens avec* $\rho = 0{,}90$.

1.3 Conditions d'ergodicité[3]

Il s'avère que le régime stationnaire existe sous les conditions usuelles, qui assurent que le flux entrant dans chaque lien ℓ est inférieur à C_ℓ.

Théorème 1.3.1. *Sous les politiques d'allocation ζ^{eq} et $\zeta^{\min}$, le réseau est*

(i) ergodique si $\max\limits_{\ell\in\mathcal{L}} \rho_\ell < 1$ *;*

(ii) transient si $\max\limits_{\ell\in\mathcal{L}} \rho_\ell > 1$.

Ce résultat a déjà été prouvé pour la politique max-min dans [17]. Toutefois, la preuve ci-dessous est plus simple et n'utilise que des résultats élémentaires sur les fonctions de Lyapounov. Elle devrait donc être plus facile a adapter à des situations plus compliquées.

Preuve. La preuve de la condition suffisante est faite pour la politique $\zeta^{\min}$, mais s'étend immédiatement à ζ^{eq} grâce au lemme 1.1.1. Quant à la preuve de la condition nécessaire, elle est valable pour une politique d'allocation de bande passante absolument quelconque.

On considère la chaîne de Markov à temps discret $(\hat{X}(n), n \in \mathbb{N})$, qui décrit la séquence d'états visités par le processus à temps continu X. Les transitions, partant d'un état $\hat{X} = \big(\hat{x}_r, r \in \mathcal{R}\big)$ fixé, sont définies comme suit :

$$\begin{aligned}
\mathbb{P}\big[\Delta\, \hat{x}_r(n) = 1 \;\big|\; \hat{X}(n) = \hat{X}\big] &= \frac{\lambda_r}{D}\,, \\
\mathbb{P}\big[\Delta\, \hat{x}_r(n) = -1 \;\big|\; \hat{X}(n) = \hat{X}\big] &= \frac{1}{D}\,\frac{\hat{x}_r}{\sigma_r}\min_{\ell\in r}\frac{C_\ell}{\hat{X}_\ell},
\end{aligned}$$

où $\Delta\, \hat{x}_r(n) \stackrel{\text{def}}{=} \hat{x}_r(n+1) - \hat{x}_r(n)$ et

$$D \stackrel{\text{def}}{=} \sum_{r\in\mathcal{R}} \Big(\lambda_r + \frac{\hat{x}_r}{\sigma_r}\min_{\ell\in r}\frac{C_\ell}{\hat{X}_\ell}\Big) \;\le\; |\mathcal{R}|\cdot\Big(\max_{r\in\mathcal{R}}\lambda_r + \max_{r\in\mathcal{R}}\frac{1}{\sigma_r}\cdot\max_{\ell\in\mathcal{L}} C_\ell\Big) \stackrel{\text{def}}{=} D'.$$

L'ergodicité du processus X est impliquée par celle de $\hat{X}$, puisque les taux de sortie d'un état X sont minorés uniformément par une constante positive.

Condition suffisante Supposons $\rho_M \stackrel{\text{def}}{=} \max_{\ell\in\mathcal{L}} \rho_\ell < 1$. On introduit la fonction de Lyapounov

$$f(\hat{X}) \stackrel{\text{def}}{=} \sum_{r\in\mathcal{R}}\ \sum_{1\le k\le \hat{x}_r} \gamma_r^k,$$

où les $\gamma_r > 1$ seront explicités plus loin. La forme de cette fonction, qui peut paraître compliquée au premier abord, a été choisie pour simplifier les calculs. On remarquera qu'elle a de fait la forme générale $\sum_{r\in\mathcal{R}} \beta_r \gamma_r^{\hat{x}_r} + K$, pour des constantes β_r et K adéquates.

3. Auteurs : Th. Deneux, G. Fayolle et J.-M. Lasgouttes

Afin d'exprimer les taux de transitions en fonction des $\hat{x}_r$ et des ρ_ℓ, on écrit

$$\hat{X}_\ell = \sum_{r \ni \ell} \hat{x}_r = \sum_{r \ni \ell} \lambda_r \sigma_r \frac{\hat{x}_r}{\lambda_r \sigma_r} \leq \rho_\ell C_\ell \max_{r \ni \ell} \frac{\hat{x}_r}{\lambda_r \sigma_r}.$$

Introduisant ensuite les quantités

$$\hat{x}_r^* \stackrel{\text{def}}{=} \frac{\hat{x}_r}{\lambda_r \sigma_r}, \quad \hat{x}_M^* \stackrel{\text{def}}{=} \max_{r \in \mathcal{R}} \hat{x}_r^*,$$

on obtient la minoration, valable $\forall r \in \mathcal{R}$,

$$\mathbb{P}\left[\Delta\, \hat{x}_r(n) = -1 \;\middle|\; \hat{X}(n) = \hat{X}\right] \geq \frac{\lambda_r}{D\rho_M} \frac{\hat{x}_r^*}{\hat{x}_M^*}.$$

On peut donc écrire :

$$\begin{aligned}
&\mathbb{E}\left[f(\hat{X}(n+1)) - f(\hat{X}(n)) \;\middle|\; \hat{X}(n) = \hat{X}\right] \\
&\quad = \sum_{r \in \mathcal{R}} \Big(\gamma_r^{\hat{x}_r + 1} \mathbb{P}\left[\Delta\, \hat{x}_r(n) = 1 \;\middle|\; \hat{X}(n) = \hat{X}\right] \\
&\qquad\qquad - \gamma^{\hat{x}_r} \mathbb{P}\left[\Delta\, \hat{x}_r(n) = -1 \;\middle|\; \hat{X}(n) = \hat{X}\right] \Big) \\
&\quad \leq \sum_{r \in \mathcal{R}} \frac{\lambda_r}{\rho_M D} \gamma_r^{\hat{x}_r} \left[\rho_M \gamma_r - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right]
\end{aligned}$$

Soit θ fixé, $\rho_M < \theta < 1$. On pose $\gamma_r \stackrel{\text{def}}{=} \gamma^{\frac{1}{\lambda_r \sigma_r}}$, où γ est choisi tel que

$$\rho_M \gamma_r = \rho_M \gamma^{\frac{1}{\lambda_r \sigma_r}} < \theta < 1, \; r \in \mathcal{R}.$$

Il vient alors, γ et θ satisfaisant les contraintes précédentes,

$$\mathbb{E}\left[f(\hat{X}(n+1)) - f(\hat{X}(n)) \;\middle|\; \hat{X}(n) = \hat{X}\right] \leq \sum_{r \in \mathcal{R}} \frac{\lambda_r \gamma^{\hat{x}_r^*}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right].$$

Soit α un réel tel que $\theta < \alpha < 1$. Les quantités suivantes seront évaluées séparément :

$$\begin{cases}
\Sigma_1 \stackrel{\text{def}}{=} \displaystyle\sum_{r : \hat{x}_r^* > \alpha \hat{x}_M^*} \frac{\lambda_r \gamma^{\hat{x}_r^*}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right], \\
\Sigma_2 \stackrel{\text{def}}{=} \displaystyle\sum_{r : \hat{x}_r^* \leq \alpha \hat{x}_M^*} \frac{\lambda_r \gamma^{\hat{x}_r^*}}{\rho_M D} \left[\theta - \frac{\hat{x}_r^*}{\hat{x}_M^*} \right].
\end{cases}$$

On majore la somme Σ_1, dont tous les termes sont négatifs, par un seul terme, correspondant précisément à un r_0 pour lequel $\hat{x}_{r_0}^* = \hat{x}_M^*$:

$$\Sigma_1 \leq \frac{\lambda_{r_0} \gamma^{\hat{x}_M^*}}{\rho_M D} (\theta - \alpha) \leq \frac{\gamma^{\hat{x}_M^*}}{\rho_M D} (\theta - \alpha) \min_{r \in \mathcal{R}} \lambda_r < 0.$$

Quant à Σ_2, on a

$$\Sigma_2 \leq \sum_{r:\hat{x}_r^* \leq \alpha \hat{x}_M^*} \frac{\lambda_r \gamma^{\alpha \hat{x}_M^*}}{\rho_M D} \theta \;\leq\; \frac{\gamma^{\hat{x}_M^*}}{\rho_M D} \gamma^{(\alpha-1)\hat{x}_M^*} |\mathcal{R}| \theta \max_{r \in \mathcal{R}} \lambda_r .$$

Finalement, si $C > 0$ et $\epsilon > 0$ sont choisis de façon à réaliser l'inégalité

$$(\theta - \alpha) \min_{r \in \mathcal{R}} \lambda_r + \gamma^{(\alpha-1)C} |\mathcal{R}| \theta \max_{r \in \mathcal{R}} \lambda_r \;\leq\; -\epsilon,$$

on obtient, $\forall \hat{X} \in \{\hat{x}_M^* > C\}$,

$$\begin{aligned} &\mathbb{E}\big[f(\hat{X}(n+1)) - f(\hat{X}(n)) \;\big|\; \hat{X}(n) = \hat{X}\big] \\ &\qquad = \;\; \Sigma_1 + \Sigma_2 \;\leq\; -\epsilon \frac{\gamma^{\hat{x}_M^*}}{\rho_M D} \;\leq\; -\epsilon \frac{\gamma^C}{\rho_M D'} \;<\; 0. \end{aligned}$$

Puisque l'ensemble $\{c_M^* \leq C\}$ est compact, on peut appliquer le théorème de Foster (voir par exemple [9]) et la chaîne est ergodique.

Condition nécessaire Supposons maintenant qu'il existe ℓ_0 tel que $\rho_{\ell_0} > 1$. En posant

$$g(\hat{X}) \stackrel{\text{def}}{=} \sum_{r \ni \ell_0} \sigma_r \hat{x}_r ,$$

on a immédiatement

$$\begin{aligned} &\mathbb{E}\big[g(\hat{X}(n+1)) - g(\hat{X}(n)) \;\big|\; \hat{X}(n) = \hat{X}\big] \\ &\qquad = \;\; \sum_{r \ni \ell_0} \sigma_r \Big(\mathbb{P}\big[\Delta\, \hat{x}_r(n) = 1 \;\big|\; \hat{X}(n) = \hat{X}\big] \\ &\qquad\qquad\qquad - \mathbb{P}\big[\Delta\, \hat{x}_r(n) = -1 \;\big|\; \hat{X}(n) = \hat{X}\big] \Big) \\ &\qquad \geq \;\; \frac{1}{D'} [C_{\ell_0} \rho_{\ell_0} - C_{\ell_0}] \;>\; 0. \end{aligned}$$

Donc, comme les sauts sont bornés, le système est transient. ■

1.4 Grandes déviations pour un réseau en étoile non symétrique [4]

La théorie des grandes déviations étudie les événements rares [5]. Ce n'est pas tant la décroissance exponentielle des probabilités qui nous intéresse ici que la manière dont ces événements rares

4. Auteurs : F. Delcoigne et A. de La Fortelle
5. Ce sont les « grandes » déviations par rapport à la moyenne, par oppositions aux « petites » déviations que décrit le théorème central limite.

se produisent. Dans le cadre des grandes déviations on peut en effet décrire la saturation du réseau, d'un lien ou d'une route. Les équations qui décrivent cette saturation « optimale » ne sont pas explicites mais algorithmiquement assez rapides à résoudre. La solution permet de mettre en évidence le ou les liens responsables de la saturation. Le même type d'équations est vérifié dans des réseaux plus généraux (notamment sous la politique max-min), qui peuvent donc être analysés de la même manière.

Considérons un réseau de N liens, chacun connecté aux $N-1$ autres, par exemple via un « hub »: il existe donc $N(N-1)/2$ routes possibles, qui sont toutes de longueur 2. Chaque route est notée $\ell\ell'$ ou $\ell'\ell$ indifféremment[6] avec $\ell, \ell' \in \mathcal{L}$.

Nous notons $\mathcal{R}$ l'ensemble des routes actives, celles pour lesquelles $\lambda_{\ell\ell'} > 0$. Nous supposons que toutes les autres routes sont initialement vides, de telle sorte que les routes $\ell\ell' \notin \mathcal{R}$ ne jouent aucun rôle.

Pour simplifier les notations, on pose $\mu_{\ell\ell'} \stackrel{\text{def}}{=} 1/\sigma_{\ell\ell'}$ et

$$\mu_{\ell\ell'}(x) \stackrel{\text{def}}{=} \mu_{\ell\ell'}\zeta_{\ell\ell'}(x) = \frac{1}{\sigma_{\ell\ell'}} \min\Big(\frac{C_\ell}{X_\ell}, \frac{C_{\ell'}}{X_{\ell'}}\Big).$$

Les résultats théoriques de cette section sont démontrés dans le rapport de recherche [3].

1.4.1 Saturation du réseau

Nous allons maintenant décrire les diverses manières de saturer le réseau. On peut faire saturer certaines routes et maintenir les autres à 0, ce qui conduit à la notion de face et à la partition de la figure 1.4.1.

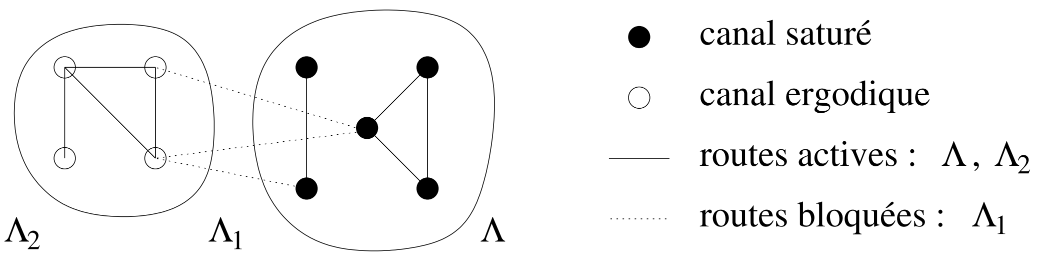

FIG. 1.4.1: *Représentation du réseau en étoile : les lignes symbolisent des routes et les cercles des liens. Les routes se partitionnent en trois classes : celles qui sont saturées (Λ), celles qui sont bloquées (Λ_1) — le taux de service y étant nul — et les routes ergodiques (Λ_2).*

Définition 1.4.1 (Face). *Soit $x \in \mathbb{R}_+^{\mathcal{R}}$, on définit la face $\Lambda(x)$ par :*

$$\Lambda(x) \stackrel{\text{def}}{=} \{\ell\ell' \in \mathcal{R} : x_{\ell\ell'} > 0\}.$$

6. Les routes ne sont donc pas orientées, mais on obtiendrait des résultats tout à fait similaires avec des routes orientées.

Par abus de notation, on appelle aussi face Λ *le sous-ensemble de* $\mathbb{R}_+^{\mathcal{R}}$ *:*

$$\left\{y \in \mathbb{R}_+^{\mathcal{R}} : y_{\ell\ell'} > 0,\ \forall \ell\ell' \in \Lambda, \quad et \quad y_{\ell\ell'} = 0,\ \forall \ell\ell' \in \Lambda^c\right\}. \tag{1.4.1}$$

À toute face Λ*, on associe une partition des routes (voir figure 1.4.1) en posant*

$$\begin{aligned} \Lambda_1 &\stackrel{\text{def}}{=} \{\ell\ell' \in \Lambda^c : \exists k \in \mathcal{L},\ \ell k \in \Lambda \text{ ou } \ell' k \in \Lambda\}, \\ \Lambda_2 &\stackrel{\text{def}}{=} \{\ell\ell' \in \Lambda^c : \forall k \in \mathcal{L},\ \ell k \notin \Lambda \text{ et } \ell' k \notin \Lambda\}. \end{aligned}$$

On définit également l'espace vectoriel relatif à Λ *:*

$$\mathbb{R}^\Lambda \stackrel{\text{def}}{=} \left\{y \in \mathbb{R}^{\mathcal{R}} : y_{\ell\ell'} = 0,\ \forall \ell\ell' \in \Lambda^c\right\}.$$

Dans un premier temps, on prouve le résultat suivant.

Théorème 1.4.2. *Si* X *est ergodique, alors, pour tout* $x \in \mathbb{R}_+^{\mathcal{R}}$ *et* $D \in \mathbb{R}^{\Lambda(x)}$*,* X *vérifie des bornes locales de grandes déviations et la fonctionnelle locale est* $L(x, D)$ *définie par*

$$L(x, D) \stackrel{\text{def}}{=} \sum_{\ell\ell' \in \Lambda(x) \cup \Lambda_1(x)} l\big(D_{\ell\ell'} \| \lambda_{\ell\ell'}, \mu_{\ell\ell'}(x)\big), \tag{1.4.2}$$

avec

$$l(D\|\lambda, \mu) \stackrel{\text{def}}{=} D \log\left(\frac{D + \sqrt{D^2 + 4\lambda\mu}}{2\lambda}\right) + \lambda + \mu - \sqrt{D^2 + 4\lambda\mu} \geq 0. \tag{1.4.3}$$

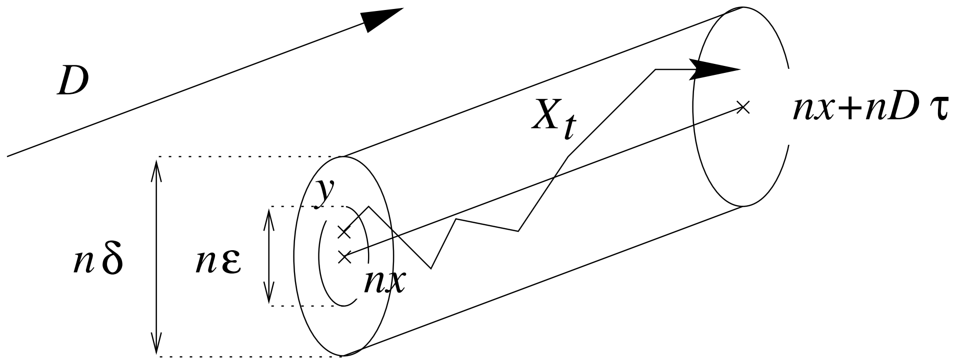

FIG. 1.4.2: *Structure des bornes linéaires locales du théorème 1.4.2.* $L(x, D)$ *est le coût par unité de temps pour que la trajectoire* $X(t, y)$ *(commençant en* y*, près de* nx*) reste au voisinage de* $nx + Dt$ *sur une durée* $t \in [0, n\tau]$.

Interprétation La structure des bornes locales de grandes déviations est illustrée par la figure 1.4.2. Il s'agit, après un changement d'échelle, de calculer au voisinage d'un point $x \in \mathbb{R}_+^{\mathcal{R}}$ la probabilité pour que le système évolue dans une direction donnée D. L'état nx représente l'état des connexions, et est choisi très grand (n tend vers l'infini), ce qui permet d'utiliser des théorèmes de limite fluide (c-à-d la loi des grands nombres).

La fonction l (1.4.3) est le « coût » par unité de temps (au sens des grandes déviations) pour qu'une M/M/1 saturée[7] de paramètres λ et μ suive le drift D. Par exemple $l(\lambda - \mu \| \lambda, \mu) = 0$, ce qui est logique puisque $D = \lambda - \mu$ est le drift.

La forme (1.4.2) de la fonctionnelle $L(x, D)$ s'interprète aisément. Localement, le réseau se comporte comme un ensemble de M/M/1 *indépendantes* de paramètres $\lambda_{\ell\ell'}$ et $\mu_{\ell\ell'}(x)$. L'indépendance est un phénomène *asymptotique et local.* Asymptotique, car il n'apparaît qu'à la limite, sur le processus renormalisé. Local, car il provient de ce que $\mu_{\ell\ell'}(.)$ reste constant au voisinage de x (c'est la continuité de $\mu_{\ell\ell'}(.)$ en x pour $\ell\ell' \in \Lambda(x) \cup \Lambda_1(x)$), et donc les variations de la position n'ont aucune incidence sur la dynamique : en fait, non seulement les routes sont indépendantes entre elles, mais la dynamique de la route $\ell\ell'$ ne dépend pas de $x_{\ell\ell'}(t)$: il s'agit de marches *homogènes.*

Ces considérations ne s'appliquent pas pour les routes $\ell\ell' \in \Lambda_2(x)$, qui restent au voisinage de 0 (où μ est discontinue). Cependant, le réseau est ergodique dans Λ_2, et le « coût » pour maintenir un réseau ergodique à 0 est nul : il n'y a donc pas de termes en $\ell\ell' \in \Lambda_2(x)$ dans l'expression de la fonctionnelle locale $L(x, D)$.

Remarquons finalement que des simplifications apparaissent : dans Λ_1, le drift est nul (on maintient les routes à 0) ainsi que $\mu_{\ell\ell'}(x)$, et $l(0 \| \lambda, 0) = \lambda$.

1.4.2 Théorème de grandes déviations

En « collant » les bornes locales de façon infinitésimale, on obtient la fonctionnelle d'action sur les trajectoires :

$$I_T(\varphi) \stackrel{\text{def}}{=} \begin{cases} \displaystyle\int_0^T L\left(\varphi(t), \dot{\varphi}(t)\right) dt, & \text{si } \varphi \text{ est absolument continue,} \\ +\infty, & \text{sinon.} \end{cases} \tag{1.4.4}$$

Une trajectoire φ est une fonction de $[0, T]$ dans $\mathbb{R}_+^{\mathcal{R}}$ continue à droite et limitée à gauche. L'espace de ces fonctions est noté $\mathbb{D}\big([0, T], \mathbb{R}_+^{\mathcal{R}}\big)$ et on le munit de la distance de Skorohod.

Remarquons que I_T est définie par les valeurs $L(x, D)$ pour $x \in \mathbb{R}_+^{\mathcal{R}}$ et $D \in \mathbb{R}^{\Lambda(x)}$ (et non pas $D \in \mathbb{R}^{\mathcal{R}}$). En effet, supposons que $\varphi_{\ell\ell'}(t) = 0$ et $\dot{\varphi}_{\ell\ell'}(t)$ existe, pour un certain t. Comme $\varphi_{\ell\ell'}(t) = 0$ est minimale ($\varphi_{\ell\ell'}$ est *positive*), nécessairement, $\dot{\varphi}_{\ell\ell'}(t) = 0$. De plus, φ est absolument continue, donc $\dot{\varphi}_{\ell\ell'}(t)$ existe pour presque tout t.

On introduit maintenant le processus renormalisé

$$X_x^n \stackrel{\text{def}}{=} \left\{ \frac{1}{n} X(nt, [nx]),\ t \geq 0 \right\}.$$

Alors le théorème de grandes déviations s'énonce comme suit.

Théorème 1.4.3. *Si X est ergodique, la suite $\{X_x^n,\ n \geq 1\}$ vérifie un principe de grandes déviations trajectoriel dans $\mathbb{D}\left([0, T], \mathbb{R}_+^{\mathcal{R}}\right)$ avec la* bonne *fonctionnelle d'action I_T, uniformément pour x appartenant à un compact.*

7. Ou, de manière équivalente, une marche aléatoire dans $\mathbb{Z}$.

Cet énoncé permet de calculer des événements *rares*, typiquement pour un ensemble $O \subset \mathbb{D}\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$ (suffisamment régulier) de trajectoires, et n assez grand,

$$\mathbb{P}\left[X^n \in O\right] \simeq \exp\left\{-n \inf_{\phi \in O} I_T(\phi)\right\}.$$

Un corollaire de ce théorème est que la probabilité stationnaire π vérifie

$$\lim_{n\to\infty} \frac{1}{n} \log \pi\big[X = [nx]\big] = -\inf_{T,\phi}\big\{I_T(\phi), \quad \phi(0) = 0,\ \phi(T) = x\big\}. \tag{1.4.5}$$

1.4.3 Saturation optimale

L'optimisation dans (1.4.5) est très délicate, compte tenu des discontinuités de $L(x, D)$ suivant les faces (et par conséquent la discontinuité de I_T). Cependant, il semble que l'on puisse conduire cette optimisation et exhiber la trajectoire optimale pour saturer un ensemble de routes (en partant d'un système vide). Les considérations suivantes reposent sur une hypothèse technique qui est vérifiée pour un certain nombre de systèmes (e.g. réseaux de Jackson, « processor sharing »). Sous cette hypothèse, on trouve que la queue de distribution pour le lien ℓ est

$$\lim_{n\to\infty} \frac{1}{n} \log \pi\big[X_\ell = n\big] = -\alpha_\ell, \tag{1.4.6}$$

où α_ℓ est l'unique solution du système en (α_ℓ, x)

$$x_{kk'} = \begin{cases} \lambda_{kk'} e^{\alpha_\ell} - \mu_{kk'}(x) e^{-\alpha_\ell} & \text{si } k = \ell \text{ ou } k' = \ell, \\ \lambda_{kk'} - \mu_{kk'}(x) & \text{si } k \neq \ell \text{ et } k' \neq \ell. \end{cases} \tag{1.4.7}$$

$$\sum_{\ell' \in \mathcal{L}} \lambda_{\ell\ell'}(1 - e^{\alpha_\ell}) + \mu_{\ell\ell'}(x)(1 - e^{-\alpha_\ell}) = 0. \tag{1.4.8}$$

Le vecteur x (dépendant de ℓ) donné par (1.4.7) décrit la « meilleure » façon de saturer le lien ℓ.

Interprétation Les équations (1.4.7)–(1.4.8) possèdent une forme assez classique, pour un œil habitué aux grandes déviations. Si l'on fixe α_ℓ, le système (1.4.7) décrit les limites fluides du réseau. Il possède quelques bonnes propriétés dans le cas de la politique min. En premier lieu, les coordonnées $x_{\ell\ell'}$ de la solution (qui est algorithmiquement constructible) sont simultanément toutes positives (resp. négatives), quand le système est transient (resp. ergodique). Puisque l'on cherche à le *saturer*, on ne considère que des solutions *positives*.

Ensuite, pour tout $\ell, \ell', k, k' \in \mathcal{L}$, $x_{\ell\ell'}$ est croissant avec $\lambda_{kk'}$ et décroissant avec $\mu_{kk'}$, ce qui signifie que *tout* le réseau sature plus rapidement en augmentant une arrivée *quelconque* ou en diminuant un départ *quelconque*. Or, augmenter α_ℓ est équivalent à augmenter $\lambda_{\ell\ell'}$ et diminuer $\mu_{\ell\ell'}$, pour tout ℓ', et le réseau sature de plus en plus.

La saturation optimale est donnée par (1.4.8), qui est une sorte de condition d'harmonicité (sans entrer dans les détails, c'est ce qui permet de prouver l'optimalité de la trajectoire). Finalement, cette interprétation décrit un algorithme de résolution des équations (1.4.7)–(1.4.8) : fixer $\alpha > 0$, résoudre (1.4.7) en x, calculer le membre de droite de l'équation (1.4.8), augmenter α si c'est positif, le diminuer sinon.

Montrons brièvement comment on résout récursivement (1.4.7). Il s'agit en fait de résoudre la limite fluide d'une file en *processor sharing* de paramètres $\hat\lambda_\ell$ et $\hat\mu_\ell$. Les équations sont similaires à (1.4.7) :

$$X_\ell = \hat\lambda_\ell - \hat\mu_\ell \frac{X_\ell}{x}, \quad \forall \ell \in \mathcal{L}, \qquad \text{avec} \quad x \stackrel{\text{def}}{=} \sum_{\ell\in\mathcal{L}} X_\ell.$$

Des manipulations algébriques montrent que x est la solution de l'équation implicite

$$\sum_{\ell\in\mathcal{L}} \frac{\hat\lambda_\ell}{x + \hat\mu_\ell} = 1. \tag{1.4.9}$$

On voit immédiatement que cette équation admet une unique solution positive si

$$\sum_{\ell\in\mathcal{L}} \frac{\hat\lambda_\ell}{\hat\mu_\ell} > 1,$$

ce qui est la condition de transience. Ces équations montrent que, pour résoudre (1.4.7), lorsque α est fixé, on doit trouver le plus grand des x_k tel que

$$\begin{aligned} \frac{\lambda_{ik} e^{\alpha_\ell}}{x_k + \mu_{ik} C_k e^{-\alpha_\ell}} + \sum_{k'\neq\ell} \frac{\lambda_{kk'}}{x_k + \mu_{kk'} C_k} &= 1, \qquad \text{si} \quad k \neq \ell, \\ \sum_{\ell'\in\mathcal{L}} \frac{\lambda_{\ell\ell'} e^{\alpha_\ell}}{X_\ell + \mu_{\ell\ell'} C_\ell e^{-\alpha_\ell}} &= 1, \qquad \text{si} \quad k = \ell. \end{aligned}$$

Si $k = \ell$, il n'y a pas besoin de poursuivre les calculs, et on est dans le cas du processor sharing : ce sont exactement les mêmes équations (en limite de grandes déviations). Le reste du réseau ne joue aucun rôle.

Sinon ($k \neq \ell$), nous n'avons pas trouvé de majoration intuitive : il faut calculer récursivement tous les x_k, dans l'ordre décroissant, en faisant également décroître la bande passante C_ℓ, puis injecter la solution dans (1.4.8), selon l'algorithme décrit précédemment.

Résolution pour une route En comparant les équivalents, pour une seule route $\ell\ell'$, des équations (1.4.7)–(1.4.8) avec le cas du processor sharing[8], on trouve la solution exacte :

$$\alpha_{\ell\ell'} = \max\left\{ \frac{\rho_{\ell\ell'}}{1 + \rho_{\ell\ell'} - \rho_\ell}, \frac{\rho_{\ell\ell'}}{1 + \rho_{\ell\ell'} - \rho_{\ell'}} \right\}, \tag{1.4.10}$$

en posant

$$\rho_{\ell\ell'} \stackrel{\text{def}}{=} \frac{\lambda_{\ell\ell'}}{C_\ell \mu_{\ell\ell'}} \qquad \text{et} \qquad \rho_\ell \stackrel{\text{def}}{=} \sum_{\ell'\in\mathcal{L}} \rho_{\ell\ell'}.$$

8. Encore une fois, on obtient exactement les mêmes équations.

Bornes pour un lien Les équations (1.4.7)–(1.4.8) ne sont pas explicites mais on peut en extraire des bornes. Tout d'abord, on constate que

$$\alpha_\ell \geq \rho_\ell, \tag{1.4.11}$$

ce qui signifie que la file sature toujours plus facilement qu'une file en processor sharing avec des arrivées $\lambda_{\ell\ell'}$ et des services $\mu_{\ell\ell'}C_\ell$, ce qui est *a priori* évident. C'est une assez bonne borne, puisque c'est la solution exacte pourvu que ce soit le lien ℓ qui sature le plus vite[9].

Trouver des bornes supérieures est assez facile, mais se complique dès que l'on cherche à les affiner. Dans un premier temps on peut vérifier que l'on a

$$\alpha_\ell \leq \max_{\ell' \in \mathcal{L}} \rho_{\ell'}, \tag{1.4.12}$$

ce qui signifie qu'on ne fait jamais moins bien que le plus mauvais lien. Cette borne n'est pas bonne dans la mesure où, à part le cas trivial $\alpha_\ell = \rho_\ell$ (il n'y a pas d'« effet réseau »), cette borne n'est jamais atteinte.

Une *minoration* facile est de borner par une quelconque des routes, grâce à l'équation (1.4.10), ce qui donne

$$\alpha_\ell \geq \max_{\ell' \in \mathcal{L}} \alpha_{\ell\ell'} = \max_{\ell' \in \mathcal{L}} \left\{ \frac{\rho_{\ell\ell'}}{1 + \rho_{\ell\ell'} - \rho_\ell}, \frac{\rho_{\ell\ell'}}{1 + \rho_{\ell\ell'} - \rho_{\ell'}} \right\}. \tag{1.4.13}$$

On vérifie que la borne (1.4.13) est toujours inférieure à (1.4.12). D'autre part, il arrive que la borne (1.4.13) soit supérieure à ρ_ℓ, auquel cas on est sûr qu'il y a un effet de réseau : le lien ℓ sature à cause d'un autre.

Extensions Au vu des techniques employées, le système d'équations (1.4.7)–(1.4.8) semble pouvoir s'étendre à des topologies et à des politiques plus générales, incluant max-min équité. Même si la solution n'a pas de forme fermée, ces équations sont algorithmiquement relativement simples à résoudre. En particulier, (1.4.10)–(1.4.13) sont encore valables pour la politique max-min.

9. C'est-à-dire que $X_\ell = \max_{\ell' \in \mathcal{L}} X_{\ell'}$ dans la solution (α_ℓ, x) de (1.4.7)–(1.4.8). C'est notamment toujours le cas dans un système symétrique.

2 Études asymptotiques

Schématiquement, il est loisible d'affirmer que la résolution de modèles raisonnablement réalistes (ce vocable étant bien sûr subjectif !) n'est possible que pour les dimensions extrêmes : faible (disons ≤ 3) ou très grande. Dans ce dernier cas, on parle de systèmes en *limite thermodynamique*, i.e. dont la taille (volume, graphe associé, nombre de nœuds, etc.), représentée par un paramètre N, augmente indéfiniment. S'il y a clairement une analogie avec les systèmes de particules rencontrés en physique, de nouvelles difficultés surgissent par suite de la non homogénéité des composants (sites).

Un des phénomènes essentiels mis en évidence est la *propagation du chaos*. Elle existe dans un réseau si, par définition, tout p-uplet de nœuds se comporte, lorsque $N \to \infty$, comme un ensemble de p nœuds indépendants : on peut alors obtenir des renseignements quantitatifs sur la dynamique, l'état stationnaire et les vitesses de convergence. Pour les systèmes à fort degré de symétrie, les techniques utilisées sont issues de la mécanique statistique et dites « à champ moyen » : asymptotiquement, chaque site aura tendance à évoluer comme un processus de saut Markovien non homogène en temps, dont les taux de transition sont déterminés par calcul de la mesure empirique des actions dues aux autres sites, laquelle devient en fait déterministe. Une des difficultés provient de la rencontre d'équations différentielles non linéaires. Nous avons démontré la propagation du chaos pour diverses classes de réseaux (BCMP [7], polling [4]). Quand la dissymétrie est totale, la plupart des problèmes restent encore largement ouverts. Or cette hypothèse est souvent nécessaire en pratique, notamment pour étudier les problèmes de goulot d'étranglement.

Même si la politique min permet d'obtenir les conditions d'ergodicité, elle reste encore trop compliquée pour permettre une résolution explicite des équations de mesure invariante. On montre donc dans ce chapitre comment l'hypothèse de propagation du chaos conduit à des équations relativement exploitables. Nous considérerons cette hypothèse comme une heuristique. On pourra trouver des études rigoureuses de ce type dans Vvedenskaya *et al.* [18] et Delcoigne et Fayolle [4]. La section 2.1 donne les équations de champ moyen de base pour notre modèle, dans un cas présentant de fortes propriétés de symétrie ; elles sont étudiées numériquement en section 2.2, puis comparées aux simulations de la section 1.2. La section 2.3 est consacrée à une étude des équations, à l'aide de divers d'outils de la variable complexe. On examine aussi comment l'analyse peut être étendue à un modèle présentant une asymétrie en section 2.4. Enfin, on montre en section 2.5 que ces équations sont en fait approchées (bien que précises) et qu'il est possible d'en obtenir de meilleures, au prix, naturellement, d'une plus grande complexité analytique.

Enfin, la section 2.6, quelque peu à part, propose une approche originale, basée sur l'étude des générateurs infinitésimaux des processus et qui permet d'étudier des modèles assez généraux, sans propriété de symétrie particulière.

2.1 Une heuristique pour des grands réseaux symétriques[1]

2.1.1 Le modèle symétrique

Chaque réseau est identifié par la valeur de N et les grandeurs qui dépendent de N sont indiquées par un exposant (N). On suppose que le réseau est symétrique, c'est-à-dire que la topologie vue de chacun des liens est la même, et que les routes sont indiscernables et toutes de même longueur.

En particulier, si $R^{(N)}$ est nombre de routes passant par un lien quelconque, alors, $\forall r \in \mathcal{R}$,

$$L_r = L > 1,\ \lambda_r^{(N)} = \lambda / R^{(N)},\ \sigma_r = \sigma.$$

Pour tout lien $\ell \in \mathcal{L}$, la capacité est fixée à $C_\ell = 1$ et la charge est donc $\rho_\ell = \rho \stackrel{\text{def}}{=} \lambda\sigma$.

Par exemple, dans le cas du réseau en étoile de la figure 1.1.1, N est pair (puisque les liens sont dédoublés), et le nombre de routes passant par un lien est $R^{(N)} = N/2$.

Il est intéressant de disposer de modèles rendant compte d'une réalité plus complexe, en particulier comportant des routes de longueur $L > 2$. La construction de tels modèles est bien sur plus difficile, notamment pour tenir compte des contraintes de symétrie.

Il s'avère qu'un modèle raisonnable est un hypercube (figure 2.1.1), dont chacune des arêtes comporte deux liens unidirectionnels.

L'hypercube est un réseau très classique, possédant de nombreuses symétries et sur lequel les calculs se conduisent bien. Il est caractérisé par sa dimension d. Ses sommets sont les d–uplets de 0 ou de 1 (ex. $(0, 1, 1, 0, 1)$) et le graphe de l'hypercube connecte tous les sommets dont une seule coordonnée diffère.

Le nombre total de liens de ce réseau est $N = d2^d$ et le nombre de routes passant par un lien est

$$R^{(N)} = L\frac{(d-1)!}{(d-L)!},$$

en se restreignant aux routes optimales dans le réseau (c'est-à-dire réalisant le plus court chemin entre deux nœuds).

Notons toutefois que les calculs présentés dans la section suivante ne dépendent pas de la topologie du réseau choisi. En particulier, tant que $R^{(N)} \to \infty$ quand $N \to \infty$, la valeur précise de $R^{(N)}$ n'apparaît pas dans les équations finales !

1. Auteurs : G. Fayolle, A. de La Fortelle et J.-M. Lasgouttes

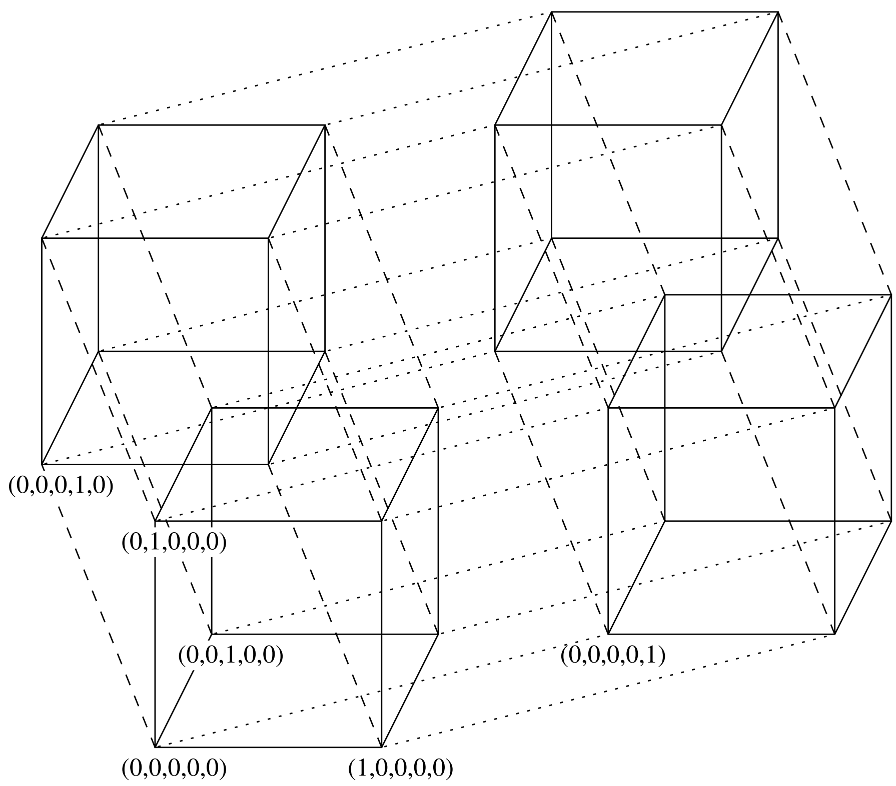

FIG. 2.1.1: *Réseau en hypercube de dimension 5 et la représentation de ses sommets.*

2.1.2 Équations de récurrence

On suppose maintenant que les conditions d'ergodicité sont satisfaites ($\rho < 1$) et que le système est dans l'état stationnaire $X^{(N)} = (x_r^{(N)}, r \in \mathcal{R})$.

Pour tout $x \geq 0$, on définit la mesure empirique donnant la proportion de liens dans l'état x,

$$\alpha_x^{(N)} \stackrel{\text{def}}{=} \frac{1}{N} \#\{\ell : X_\ell^{(N)} = x\}.$$

Par symétrie, il est clair que, pour tout $\ell \in \mathcal{L}$, $\mathbb{P}(X_\ell^{(N)} = x) = \mathbb{E}\alpha_x^{(N)}$. D'autre part, nous exprimons l'hypothèse de propagation du chaos par le fait que $\alpha_x^{(N)}$ vérifie une forme de loi des grands nombres :

$$\alpha_x \stackrel{\text{def}}{=} \lim_{N\to\infty} \alpha_x^{(N)} = \lim_{N\to\infty} \mathbb{P}(X_\ell^{(N)} = x) \stackrel{\text{def}}{=} \mathbb{P}(X = x).$$

$\alpha_x^{(N)}$ est traditionnellement appelé *champ moyen* et il gouverne la dynamique du système. On pose aussi

$$\bar{\alpha}^{(N)} \stackrel{\text{def}}{=} \mathbb{E}X_\ell^{(N)} = \sum_{x>0} x\alpha_x^{(N)}, \qquad \bar{\alpha} \stackrel{\text{def}}{=} \sum_{x>0} x\alpha_x.$$

Avant de déterminer la distribution stationnaire limite α_x il faut décrire les transitions possibles de $\alpha_x^{(N)}$. Les deux cas possibles sont :

- arrivée sur un lien avec x connexions
$$\alpha_x \to \alpha_x - \frac{1}{N}, \quad \alpha_{x+1} \to \alpha_{x+1} + \frac{1}{N}$$
- départ d'un lien avec $x > 0$ connexions
$$\alpha_x \to \alpha_x - \frac{1}{N}, \quad \alpha_{x-1} \to \alpha_{x-1} + \frac{1}{N}$$

Sur chaque lien, la transition $x \to x+1$ a lieu au taux λ, par suite de l'indépendance des arrivées externes. Pour calculer l'intensité de départ d'un lien, nous aurons besoin du lemme suivant.

Lemme 2.1.1. *Supposons que les $C^{(N)}$ connexions présentes dans le système soient réparties de manière uniforme et indépendante et que, quand $N \to \infty$, $C^{(N)}/N \to C > 0$ et $R^{(N)} \to \infty$. Pour toute route r, on a alors, lorsque $N \to \infty$:*
$$\mathbb{E}\left[x_r^{(N)} \middle| X_{r(1)}^{(N)} = y_1, \ldots, X_{r(L)}^{(N)} = y_L\right] \approx \frac{y_1 \cdots y_L}{R^{(N)}} \cdot \left(\frac{N}{LC^{(N)}}\right)^{L-1}. \tag{2.1.1}$$

Preuve. Soit r une route arbitraire, sur laquelle le nombre de connexions en cours a une valeur fixée x, avec bien sûr $x \leq y_\ell$ pour $1 \leq \ell \leq L$. Le tableau ci-dessous récapitule la combinatoires des répartitions des $C^{(N)}$ connexions vérifiant $X_{r(i)} = y_i$ (on notera que le nombre total de routes dans le réseau est $NR^{(N)}/L$) :

Ensemble	Cardinal	Nb. connexions
r	1	x
$\{r' \neq r, r(i) \in r'\}$	$R^{(N)} - 1$	$y_i - x$
Autres	$\frac{NR^{(N)}}{L} - L(R^{(N)} - 1) - 1$	$C^{(N)} - \sum_{i=1}^{L}(y_i - x) - x$

Le nombre total de répartitions possibles est alors égal à
$$f(x) \stackrel{\text{def}}{=} \prod_{i=1}^{L} \binom{R^{(N)}-2+y_i-x}{y_i-x} \cdot \binom{\frac{NR^{(N)}}{L}-L(R^{(N)}-1)-2+C^{(N)}-\sum_{i=1}^{L}(y_i-x)-x}{C^{(N)}-\sum_{i=1}^{L}(y_i-x)-x}$$

Lorsque $N \to \infty$, il vient
$$\begin{aligned}\frac{f(1)}{f(0)} &= \prod_{i=1}^{L}\left[\frac{y_i}{R^{(N)}-2+y_i-1}\right] \cdot \frac{(C^{(N)}-\sum_{i=1}^{L} y_i)!}{(C^{(N)}-\sum_{i=1}^{L} y_i + (L-1))!} \\ &\quad \cdot \frac{(\frac{NR^{(N)}}{L} - L(R^{(N)}-1) - 2 + C^{(N)} - \sum_{i=1}^{L} y_i + (L-1))!}{(\frac{NR^{(N)}}{L} - L(R^{(N)}-1) - 2 + C^{(N)} - \sum_{i=1}^{L} y_i)!} \\ &\approx \frac{\prod_{i=1}^{L} y_i}{(R^{(N)})^L} \cdot \frac{1}{(C^{(N)})^{L-1}} \cdot \left(\frac{NR^{(N)}}{L}\right)^{L-1},\end{aligned}$$

et la preuve de ce lemme est terminée. ■

Pour un lien ℓ tel que $X_\ell^{(N)} = x$, en remarquant que $C^{(N)} = N\bar{\alpha}^{(N)}/L$, le débit instantané vaut approximativement

$$
\begin{aligned}
D_x^{(N)} &\stackrel{\text{def}}{=} \frac{1}{\sigma}\,\mathbb{E}\Bigg[\sum_{r\ni\ell} \frac{x\cdot X_{r(2)}^{(N)}\cdots X_{r(L)}^{(N)}}{R^{(N)}\,(\bar{\alpha}^{(N)})^{L-1}}\left(\frac{1}{x}\wedge\frac{1}{X_{r(2)}^{(N)}}\wedge\cdots\wedge\frac{1}{X_{r(L)}^{(N)}}\right)\Bigg]\\
&= \frac{1}{\sigma(\bar{\alpha}^{(N)})^{L-1}}\mathbb{E}\sum_{y_2,\dots y_L=1}^{\infty}\left(\frac{1}{x}\wedge\frac{1}{y_2}\wedge\cdots\wedge\frac{1}{y_L}\right)y_2\alpha_{y_2}^{(N)}\cdots y_L\alpha_{y_L}^{(N)}.
\end{aligned}
$$

Prenant la limite $N\to\infty$, on en déduit les équations de mesure invariante qui s'écrivent

$$
-\lambda\alpha_0 + \frac{1}{\sigma}\alpha_1\frac{u_1}{\bar{\alpha}^{L-1}} = 0, \tag{2.1.2}
$$

$$
\lambda(\alpha_{x-1}-\alpha_x) + \frac{1}{\sigma}\left(\alpha_{x+1}\frac{u_{x+1}}{\bar{\alpha}^{L-1}} - \alpha_x\frac{u_x}{\bar{\alpha}^{L-1}}\right) = 0 \qquad \forall x\ge 1, \tag{2.1.3}
$$

où on a posé, pour $x\ge 0$,

$$
u_x \stackrel{\text{def}}{=} x\sum_{y_2,\dots y_L=1}^{\infty}\left(\frac{1}{x}\wedge\frac{1}{y_2}\wedge\cdots\wedge\frac{1}{y_L}\right)y_2\alpha_{y_2}\cdots y_L\alpha_{y_L}. \tag{2.1.4}
$$

Reportant (2.1.2) dans (2.1.3), on obtient immédiatement la relation fondamentale

$$
\alpha_{x+1}u_{x+1} = \rho\bar{\alpha}^{L-1}\alpha_x, \qquad \forall x\ge 0. \tag{2.1.5}
$$

Une caractéristique intéressante de ces équations est qu'elles ne dépendent absolument pas de la topologie du réseau considéré. Ce fait est très surprenant et on aurait pu *a priori* s'attendre à l'apparition de $R^{(N)}$, sous une forme plus une moins explicite.

Afin de récrire plus agréablement u_x, on introduit $Y_2,\dots,Y_L$, variables aléatoires i.i.d. vérifiant

$$
\mathbb{P}(Y_i = y) = \frac{y\alpha_y}{\bar{\alpha}},\ \ 2\le i\le L,
$$

et

$$
Y \stackrel{\text{def}}{=} \frac{1}{Y_2}\wedge\cdots\wedge\frac{1}{Y_L}.
$$

Alors, l'équation (2.1.4) devient

$$
u_x = x\bar{\alpha}^{L-1}\,\mathbb{E}\Big[\frac{1}{x}\wedge Y\Big]. \tag{2.1.6}
$$

Par des propriétés classiques du minimum de variables aléatoires indépendantes, pour un i quelconque,

$$
\mathbb{P}\Big(Y\ge\frac{1}{y}\Big) = \Big[\mathbb{P}\Big(\frac{1}{Y_i}\ge\frac{1}{y}\Big)\Big]^{L-1} = \big[\mathbb{P}(Y_i\le y)\big]^{L-1} = \Big[\sum_{m=0}^{y}\frac{m\alpha_m}{\bar{\alpha}}\Big]^{L-1},
$$

et donc

$$\begin{aligned}
\mathbb{E}\Big[\frac{1}{x}\wedge Y\Big] &= \sum_{y=1}^{x}\frac{1}{x}\mathbb{P}\Big(Y=\frac{1}{y}\Big)+\sum_{y=x+1}^{\infty}\frac{1}{y}\mathbb{P}\Big(Y=\frac{1}{y}\Big)\\
&= \frac{1}{x}\mathbb{P}\Big(Y\geq\frac{1}{x}\Big)+\sum_{y=x+1}^{\infty}\frac{1}{y}\Big[\mathbb{P}\Big(Y\geq\frac{1}{y}\Big)-\mathbb{P}\Big(Y\geq\frac{1}{y-1}\Big)\Big]\\
&= \Big[\frac{1}{x}-\frac{1}{x+1}\Big]\mathbb{P}\Big(Y\geq\frac{1}{x}\Big)+\sum_{y=x+1}^{\infty}\Big[\frac{1}{y}-\frac{1}{y+1}\Big]\mathbb{P}\Big(Y\geq\frac{1}{y}\Big)\\
&= \sum_{y=x}^{\infty}\frac{1}{y(y+1)}\mathbb{P}\Big(Y\geq\frac{1}{y}\Big)
\end{aligned}$$

Finalement, la forme simplifiée de u_x est :

$$u_x = x\sum_{y=x}^{\infty}\frac{1}{y(y+1)}\Big[\sum_{m=0}^{y}m\alpha_m\Big]^{L-1}.$$

On notera que dans le cas $L=2$, étudié dans la section 2.3 ci-après, la forme des équations se simplifie quelque peu :

$$\begin{aligned}
u_x &= \sum_{y>0}(x\wedge y)\alpha_y,\\
\alpha_{x+1}u_{x+1} &= \rho\bar{\alpha}\alpha_x, \qquad \forall x\geq 0.
\end{aligned} \tag{2.1.7}$$

2.2 Résolution numérique des équations[2]

Les solutions analytiques explicites de la plupart des équations étant hors de portée, il est intéressant de les étudier numériquement pour dégager l'influence des différents paramètres. La résolution numérique de l'équation (2.1.5) peut s'effectuer par une méthode de point fixe : on se donne les α_x *a priori*, on calcule les u_x, puis on en déduit les $\alpha_x^{(1)}$ vérifiant (2.1.5), et ainsi de suite. Avec quelques précautions contre les instabilités numériques, cette méthode converge assez rapidement.

La figure 2.2.1 montre une forme de distribution modale proche de celle obtenue par simulation (figure 1.2.1). Pour s'en persuader, il suffit de comparer les distributions cumulées des simulations et du champ moyen obtenu pour $L=2$ (figure 2.2.2). On voit que la courbe obtenue pour $N=300$ coïncide bien avec celle obtenue par résolution du champ moyen. Toutefois, comme on le verra en section 2.5, il n'y a pas vraiment convergence vers le champ moyen, même si l'heuristique est très bonne.

D'autre part, on note que le cas $L=20$ (figure 2.2.3) n'est pas très différent du cas ($L=2$) du réseau en étoile. En fait, en comparant les deux, on s'aperçoit que les pics sont répartis de la même manière à un facteur 5 près.

2. Auteur : A. de La Fortelle

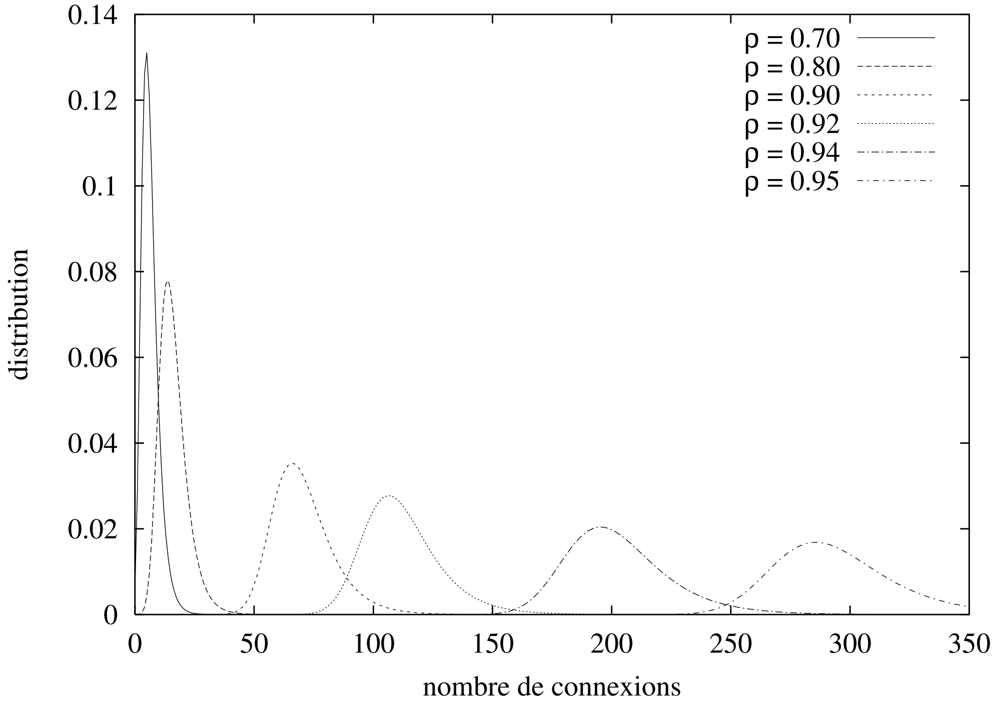

FIG. 2.2.1: *Distributions en champ moyen (cas $L = 2$, réseau en étoile), pour différentes charges du réseau.*

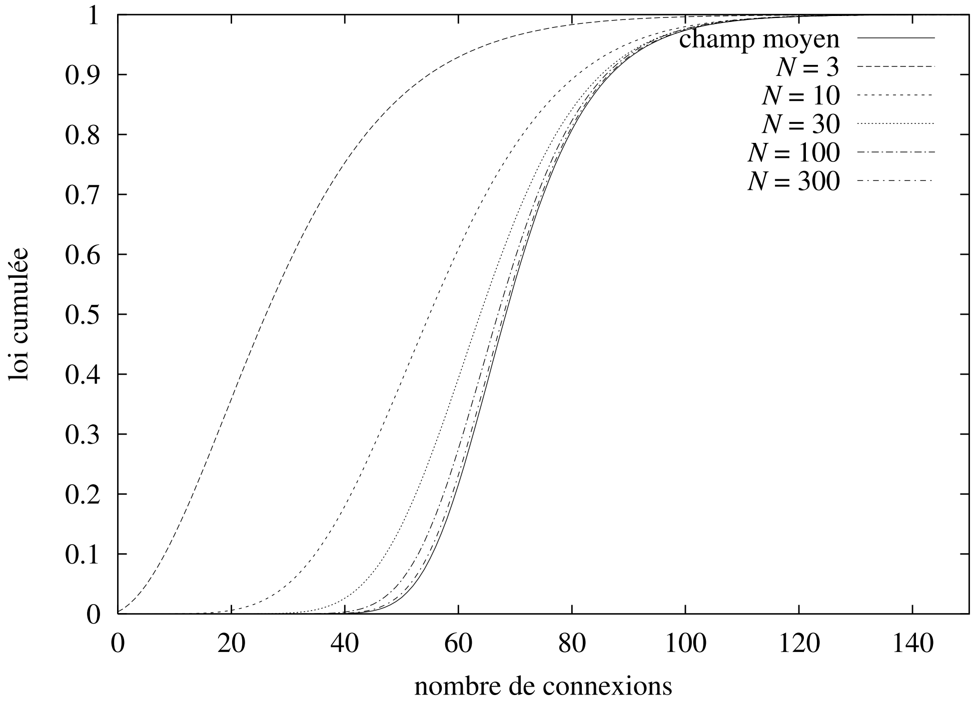

FIG. 2.2.2: *Comparaison des distributions cumulées entre le champ moyen et la simulation pour différents N (réseau en étoile, $\rho = 0{,}90$).*

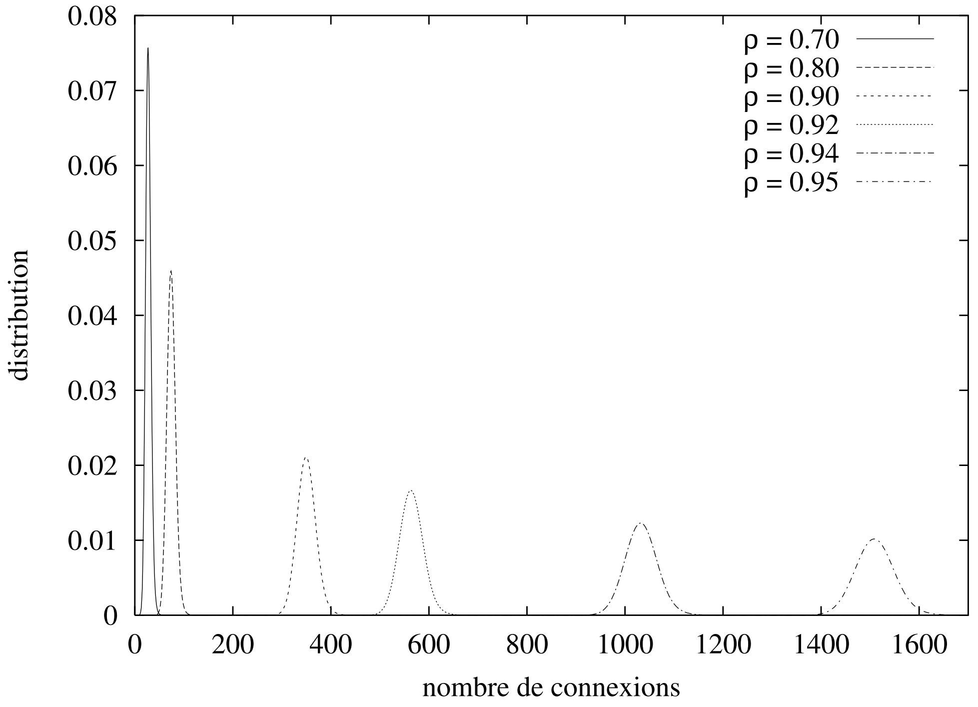

FIG. 2.2.3: *Distributions en champ moyen (cas $L = 20$), pour différentes charges du réseau.*

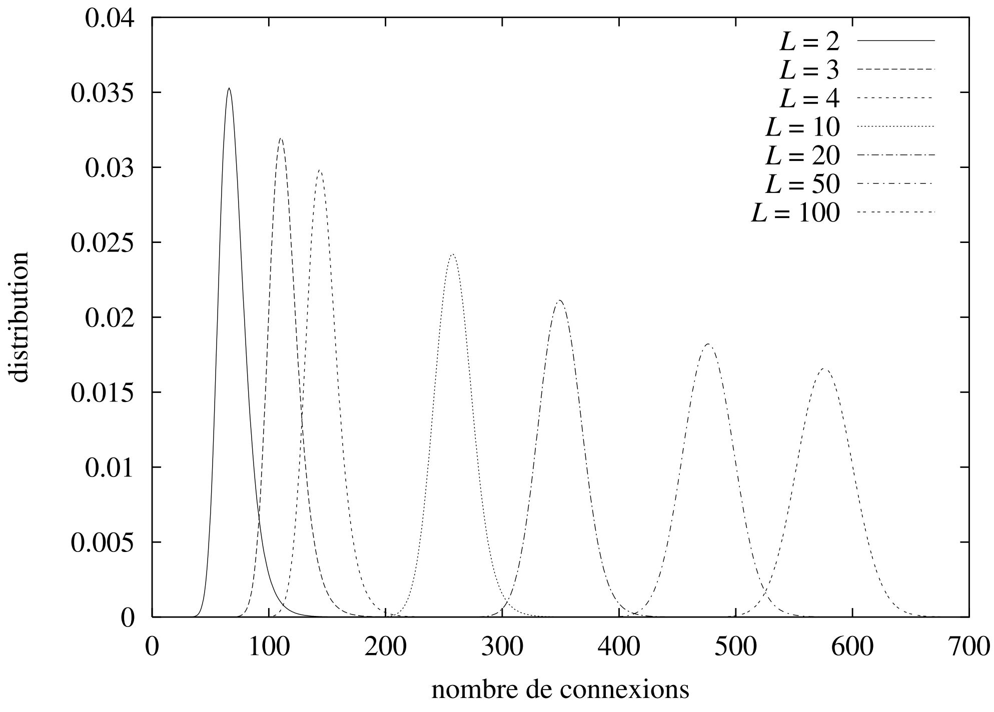

FIG. 2.2.4: *Distribution en champ moyen pour des routes de différentes longueurs et $\rho = 0{,}9$.*

La figure 2.2.4 montre l'effet d'une modification de la longueur des routes pour une même valeur de ρ. Une caractéristique surprenante de ces courbes est qu'il ne semble pas y avoir de différence d'ordre de grandeur entre les distributions quand la longueur varie. La figure 2.2.5 que la moyenne est toujours en $K/(1-\rho)^2$, ou seule la constante K varie avec la longueur : on voit bien que la partie gauche ($\rho \to 1$: charge élevée) des courbes est rectiligne, ce dont nous déduisons que le comportement de la moyenne est une puissance de $1-\rho$. S'il on calcule la pente de ces droites, on trouve justement -2. On aurait pu craindre, par exemple, que la moyenne se comporte en $(1-\rho)^{-L}$, comme la comparaison entre M/M/1 (longueur 1, moyenne en $(1-\rho)^{-1}$) et le réseau en étoile (longueur 2, moyenne en $(1-\rho)^{-2}$) semble le suggérer.

Il est intéressant enfin d'étudier l'évolution de la constante K en fonction de L. Il s'avère en fait que sa croissance est à peu près logarithmique, comme le montre la figure 2.2.6.

Ces résultats sont d'un intérêt pratique important, car ils montrent qu'il n'y a pas de phénomène d'engorgement dans les réseaux lorsque la taille des routes est grande. D'une part, le changement de comportement sous forte charge ($\rho \to 1$) de la moyenne entre $L = 1$ ($\approx (1-\rho)^{-1}$) et $L = 2$ ($\approx (1-\rho)^{-2}$) ne s'amplifie pas pour $L > 2$. L'effort d'optimisation devrait donc pouvoir être « local », dans un sens qui reste à déterminer. D'autre part, la croissance semble très lente et de type logarithmique, montrant encore une fois que la longueur des routes n'influence que faiblement le comportement du réseau. Enfin, la politique min étant majorante de max-min, ces conclusions semblent *a priori* valables pour les réseaux dont l'allocation de la bande passante est max-min équitable.

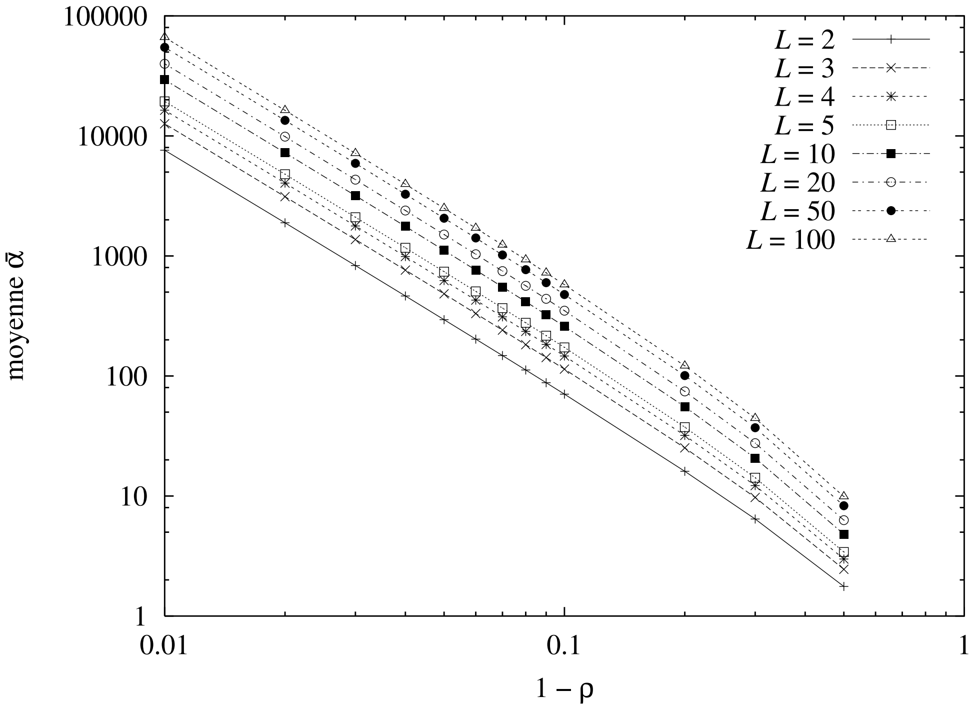

Fig. 2.2.5: *Évolution de la moyenne des distributions en fonction de la charge, pour des routes de différentes longueurs.*

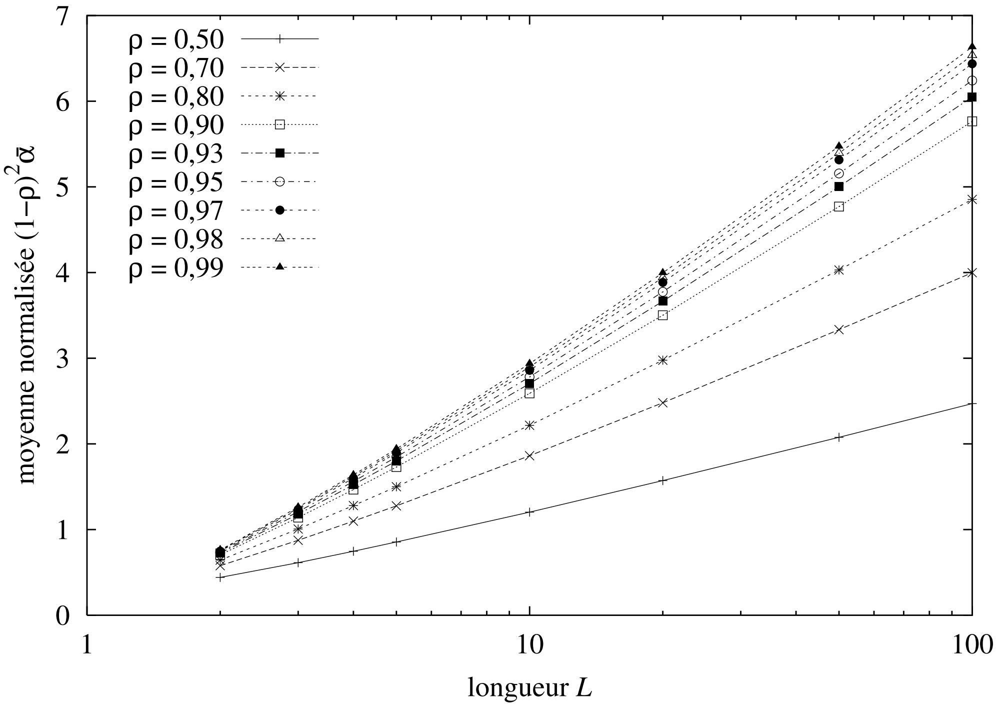

Fig. 2.2.6: *Évolution de* $(1-\rho)^2\bar{\alpha}$ *en fonction de* L *pour différentes valeurs de* ρ

2.3 Étude analytique[3]

L'évolution des α_k régie par la relation (2.1.7), en dépit d'un aspect extérieur « processus de naissance et de mort », est beaucoup plus complexe, car non linéaire et, en fait, non Markovienne, à cause de la forme de u_k et de la constante $\bar{\alpha}$. Nous proposons ci-après une méthode de résolution basée sur l'analyse fonctionnelle.

Le résultat fondamental de l'étude, contenu dans le théorème 2.3.5, a la formulation concrète suivante : lorsque $\rho \to 1$, on a les développements

$$\begin{aligned}\bar{\alpha} &\approx \frac{1}{(1-\rho)^2 A},\\ \lim_{k\to\infty} \rho^{-k}\alpha_k &\approx (1-\rho)B\exp\Big[\frac{1}{(1-\rho)A}\Big],\end{aligned}$$

où A et B sont des constantes positives avec

$$A = \int_0^\infty c(z,1)dz = \lim_{z\to\infty} zv'(z,1) \approx 1{,}30\,,$$

c et v étant les solutions du système différentiel (2.3.15). Par suite de la forte instabilité numérique de ce système, il semble difficile avec le logiciel utilisé (« Livermore stiff ODE » de MAPLE) d'obtenir une valeur sûre plus précise de la constante A.

2.3.1 Une équation intégrale pour la fonction génératrice

Soit la fonction génératrice $\alpha : z \to \alpha(z)$, a priori définie pour z dans le disque unité fermé,

$$\alpha(z) \stackrel{\text{def}}{=} \sum_{k\geq 0} \alpha_k z^k.$$

En désignant par α' la dérivée de α, on a $\alpha'(1) = \bar{\alpha}$ et la relation (2.1.7) se met sous la forme

$$\alpha_{k+1}u_{k+1} = \rho\alpha'(1)\alpha_k, \quad \forall k \geq 0. \tag{2.3.1}$$

On notera en général $\mathcal{C}(r)$ et $\mathcal{D}(r)$, respectivement, le cercle de rayon r et le disque ouvert de rayon r.

Lemme 2.3.1.

(a) Si (2.3.1) a une solution probabiliste, alors, nécessairement, $\rho < 1$ et

$$\lim_{k\to\infty} \alpha_k \rho^{-k} = K(\rho),$$

où $K(\rho)$ est une constante positive, bornée $\forall \rho < 1$.

3. Auteurs : G. Fayolle et J.-M. Lasgouttes

(b) La fonction α satisfait l'équation intégrale non linéaire

$$\alpha'(1)(1-\rho z)\alpha(z) = \frac{1}{2i\pi}\int_{\mathcal{C}(r)} \alpha(\omega)\,\alpha\Big(\frac{z}{\omega}\Big)\frac{d\omega}{(1-\omega)^2}\,, \tag{2.3.2}$$

où $|z| < \rho^{-1}$ et r est un nombre positif arbitraire, avec $1 < r < \rho^{-1}$.

Preuve. Lorsque (2.1.7) possède une solution probabiliste, on a toujours $\lim_{m\to\infty} u_m = \bar{\alpha}$. Dans ce cas, $\forall\epsilon > 0$, il existe un nombre $M(\epsilon) > 0$, tel que

$$\rho \le \frac{\alpha_{m+1}}{\alpha_m} \le \frac{\rho}{1-\epsilon}, \quad \forall m \ge M(\epsilon),$$

ce qui implique $\rho < 1$. En outre, sous la même hypothèse d'existence, on peut écrire, en posant $D_l = \sum_{m\ge l}(m-l)\alpha_m$,

$$\alpha_k\rho^{-k} = \alpha_0 \prod_{l\ge 0}^{k} \frac{\bar{\alpha}}{\bar{\alpha} - D_{l+1}}. \tag{2.3.3}$$

Lorsque $l \to \infty$, la convergence du produit dans (2.3.3) équivaut à celle de la série de terme général D_l, qui converge car

$$D_l \le \alpha_0 \sum_{m\ge l}(m-l)(\rho+\epsilon)^m,\ \forall l \ge M(\epsilon),$$

de sorte que $\sum_l D_l$ se comporte comme

$$\sum_{l\ge 0}\sum_{n\ge 0} n(\rho+\epsilon)^{l+n} = \frac{\rho+\epsilon}{(1-\rho-\epsilon)^3}.$$

Le point (a) du lemme est démontré.

Le point (b) est une application de (a) et d'une représentation intégrale utilisée par Hadamard, qu'on rappelle ci-dessous (cf. par exemple [15]).

Soit les fonctions

$$a(z) = \sum_{n\ge 0} a_n z^n, \qquad b(z) = \sum_{n\ge 0} b_n z^n,$$

analytiques respectivement dans les disques $\mathcal{D}(R)$ et $\mathcal{D}(R')$. Alors la fonction

$$c(z) = \sum_n a_n b_n z^n$$

a un rayon de convergence au moins égal à RR' et elle admet la représentation intégrale

$$c(z) = \frac{1}{2i\pi}\int_{\mathcal{L}} a(\omega)\, b\Big(\frac{z}{\omega}\Big)\frac{d\omega}{\omega}\,,$$

où $\mathcal{L}$ est un contour fermé contenant l'origine, sur lequel $|\omega| < R$, $\left|\frac{z}{\omega}\right| < R'$.

D'après (a), α a un rayon de convergence au moins égal à ρ^{-1} : cette propriété, utilisée dans la formule d'Hadamard, conduit directement à l'équation (2.3.2). ■

Le lemme suivant donne la morphologie de la fonction $\alpha(z)$.

Lemme 2.3.2. *La fonction α est méromorphe et s'écrit sous la forme*

$$\alpha(z) = \sum_{i=1}^{\infty} \frac{a_i}{\rho^{-i} - z} \tag{2.3.4}$$

et, pour toute suite de cercles $\mathcal{C}(R_n)$ tels que

$$(1+\epsilon)\rho^{-n} \leq R_n \leq (1-\epsilon)\rho^{-(n+1)}, \quad 0 < \epsilon < \frac{1-\rho}{1+\rho},$$

on a $|\alpha(R_n)| = o(1)$ lorsque $n \to \infty$.

Preuve. L'équation intégrale (2.3.2) permet de prolonger analytiquement la fonction α dans tout le plan complexe. En effet, d'après le point (a) du lemme 2.3.1, α est holomorphe dans le disque $\mathcal{D}(\rho^{-1})$ et sa première singularité est un pôle simple au point $z = \rho^{-1}$. On applique alors le théorème de Cauchy à l'intégrale dans (2.3.2), ce qui conduit à

$$\frac{1}{2i\pi} \int_{\mathcal{C}(R_1)} = \frac{1}{2i\pi} \int_{\mathcal{C}(r)} + \text{Résidu}(\rho^{-1}),$$

où R_1 est donné dans l'énoncé du lemme. Puisque $\alpha(\omega)$ et $\alpha(z/\omega)$ sont analytiques dans les régions respectives

$$|\omega| < \rho^{-1}, \quad \left|\frac{z}{\omega}\right| < \rho^{-1},$$

l'intégrale est une fonction de z analytique dans la couronne $\rho^{-1} < |z| < \rho^{-2}$, et il en est de même de α, grâce à la partie gauche de (2.3.2). La suite se fait par récurrence.

Pour établir la deuxième propriété du lemme, on applique le théorème de Cauchy en évaluant l'intégrale de la partie droite de (2.3.2) sur le cercle $\mathcal{C}(R_n)$. D'où

$$\alpha'(1)(1-\rho z)\alpha(z) - \sum_{i=1}^{n} \frac{a_i \alpha(z\rho^i)}{(\rho^{-i} - 1)^2} = \int_{\mathcal{C}(R_n)} \alpha(\omega)\, \alpha\Big(\frac{z}{\omega}\Big) \frac{d\omega}{(1-\omega)^2}.$$

Il est loisible de prolonger analytiquement l'égalité ci-dessus, dont le membre gauche est une fonction analytique dans $|z| \leq R_n$, en laissant z atteindre le cercle $\mathcal{C}(R_n)$, le long d'une courbe simple qui évite les singularités polaires $\rho^{-i}, i = 1, \ldots, n$. Finalement, la majoration du module de l'intégrale conduit à l'inégalité

$$(\rho R_n - 1)\mathcal{M}(R_n) \leq A_n \sup_{i \leq n-1} (\mathcal{M}(R_i)) + D\mathcal{M}(R_n) \int_{\mathcal{C}(R_n)} \frac{|d\omega|}{|1-\omega|^2},$$

où D est une constante positive et A_n est bornée $\forall n \leq \infty$. Par induction, on déduit alors sans difficulté $\lim_{n\to\infty} \mathcal{M}(R_n) = 0$ et la démonstration du lemme est terminée. ■

En portant la forme (2.3.4) dans (2.3.2), on obtient, par un calcul direct de résidus,

$$\alpha'(1)(1-\rho z)\sum_{i=1}^{\infty}\frac{a_i}{\rho^{-i}-z} = \sum_{i,j=1}^{\infty}\frac{a_i a_j \rho^{-j}}{(\rho^{-(i+j)}-z)(1-\rho^{-j})^2}. \tag{2.3.5}$$

Soit la fonction génératrice $a : z \to a(z)$,

$$a(z) \stackrel{\text{def}}{=} \sum_{k\geq 0} a_{k+1} z^k,$$

définie pour z dans un domaine borné du plan complexe $\mathbb{C}$, contenant l'origine. En utilisant le point (a) du lemme 2.3.1 et (2.3.4), on obtient les relations

$$\begin{cases} a_1 = \rho K(\rho) \\ \alpha_{k-1} = \rho^k a(\rho^k), \quad \forall k \geq 1, \end{cases} \tag{2.3.6}$$

qui montrent notamment, puisque $\alpha_0 < 1-\rho$, que a est analytique dans le disque $\mathcal{D}(\rho)$. L'identification des coefficients des puissances de z dans (2.3.5) permet d'écrire

$$\alpha'(1)(\rho^{k-1}-1)a_k = \sum_{j=1}^{k-1}\frac{a_j a_{k-j}\rho^{j+k-1}}{(1-\rho^j)^2}, \quad \forall k \geq 1, \tag{2.3.7}$$

et on voit aisément par récurrence que les coefficients a_i sont de signes alternés, a_1 étant positif, $\forall \rho < 1$.

Soit

$$f(t) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty}\left(\frac{\rho^j}{1-\rho^j}\right)^2 t^j = t\sum_{j=1}^{\infty}\frac{j\rho^{j+1}}{1-t\rho^{j+1}}, \quad \rho < 1,\ |t| < \rho^{-2}. \tag{2.3.8}$$

L'application de la formule d'Hadamard dans (2.3.7) permet aisément l'écriture de l'équation intégro-fonctionnelle

$$\alpha'(1)[a(\rho z) - a(z)] = \frac{a(\rho z)}{2i\pi}\int_{\mathcal{C}(r)} a(\omega) f\left(\frac{z}{\omega}\right) d\omega\,, \tag{2.3.9}$$

valable dans le domaine $\left\{r \leq \rho,\ |z| < \rho^{-2}\right\}$.

Choisissant f sous la forme de la seconde série de (2.3.8), qui converge d'ailleurs aussi le domaine $\left\{\rho < 1,\ \Re(z) \leq 0\right\}$, le théorème des résidus appliqué à l'intégrale (2.3.9) conduit à l'équation fonctionnelle

$$\alpha'(1)[a(\rho z) - a(z)] = a(\rho z) b(z), \quad |z| \leq \rho,\ \rho < 1. \tag{2.3.10}$$

avec

$$b(z) \stackrel{\text{def}}{=} z\sum_{j=1}^{\infty} j\rho^{j+1} a(\rho^{j+1} z),$$

en notant que $b(1) = \alpha'(1)$ d'après (2.3.6).

Partant de (2.3.10), on effectue le prolongement analytique de a dans les disques imbriqués $\mathcal{D}(\rho^{-n}), n \geq 1$. Il en résulte que a n'a aucune singularité à distance finie et, par conséquent, est une fonction *entière*, laquelle, d'après la théorie générale [15], est complètement caractérisée par ses zéros et son *ordre* à l'infini.

Soit z_0 un zéro arbitraire de a. Alors, d'après (2.3.10), $a(\rho^{-i}z_0) = 0$ et les zéros de a forment des familles de points en progression géométrique de raison ρ^{-1}. Il suffit donc de connaître les zéros de plus petit module : ceci n'est possible, en toute généralité, que numériquement, les difficultés de précision rencontrées étant d'ailleurs assez prohibitives !

Partant de l'identité $b(1) = \alpha'(1)$, il est utile de noter les relations

$$a(\rho^{-i}) = 0, \quad \forall i \geq 0,$$

qui seront utilisées dans la suite.

2.3.2 Sur le comportement asymptotique autour de $\rho = 1$

Pour mesurer l'intérêt pratique de la politique min, il est essentiel de comprendre le comportement du système à forte charge. Les calculs numériques de 2.2 montrent l'existence d'un *mode*, phénomène peu fréquent dans la plupart des systèmes classiques.

Au cours de cette section, il sera commode de considérer que ρ est une variable (et non plus seulement un paramètre), qu'on introduira explicitement dans les fonctions et quantités concernées, sans autre avertissement et en évitant toute ambiguïté. Par exemple, on écrira $a(z,\rho)$, $f(z,\rho)$, $a_k(\rho)$,...

Les idées fondamentales de l'analyse seront exposées après le lemme suivant, qui exhibe un changement d'échelle (peut-être le seul intéressant) relatif à la fonction $a(z,\rho)$.

Soit

$$\begin{cases} \xi(\rho) \stackrel{\text{def}}{=} \dfrac{-a_1(\rho)}{(\log\rho)^3\alpha'(1,\rho)}, \\ c_k(\rho) \stackrel{\text{def}}{=} \dfrac{-a_k(\rho)}{(\log\rho)^3\alpha'(1,\rho)\xi(\rho)^k}, \quad \forall k \geq 1, \\ c(z,\rho) \stackrel{\text{def}}{=} \displaystyle\sum_{i=0}^{\infty} c_{k+1}(\rho)z^k. \end{cases} \tag{2.3.11}$$

Le lecteur pourra se convaincre qu'en fait le facteur $(1-\rho)^3$ intervient de façon mathématiquement assez *naturelle* ; nous avons préféré $-\log^3\rho$ pour des raisons de compacité dans la présentation des formules mathématiques à venir.

Lemme 2.3.3. *Soit*

$$v(z,\rho) \stackrel{\text{def}}{=} \frac{(\log\rho)^2}{2i\pi}\int_{\mathcal{C}(1)} c(\omega,\rho) f\Big(\frac{z}{\omega},\rho\Big)d\omega, \quad |z| < \rho^{-2}.$$

On a les relations fonctionnelles suivantes :

$$\begin{cases} a(z,\rho) = a_1(\rho)\, c(z\xi(\rho),\rho), \\ c(0,\rho) = 1, \\ c(z,\rho) = c(\rho z,\rho)\big[1 + \log\rho\, v(z,\rho)\big], \\ \qquad\quad = c(\rho z,\rho)\big[1 + z(\log\rho)^3 \displaystyle\sum_{i=1}^{\infty} i\rho^{i+1} c(\rho^{i+1}z,\rho)\big]. \end{cases} \tag{2.3.12}$$

Les coefficients $c_k(\rho)$, $k \geq 1$, sont de signes alternés et la fonction $c(z,\rho)$ jouit des propriétés ci-après.

(a) Il existe une seule solution $c(z,\rho)$ de (2.3.11), fonction entière de la variable z et bi-analytique en (z,ρ) dans la région $0 < \rho < 1$.

(b) Posons

$$g(t) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \frac{t^j}{j^2} = \int_0^t \frac{-\log(1-u)du}{u}, \quad \forall |t| \leq 1,$$

que certains auteurs notent $\mathrm{dilog}(1-t)$*. Alors $c(z,1)$ existe et satisfait l'équation intégro-différentielle*

$$z\frac{\partial c(z,1)}{\partial z} = -\frac{c(z,1)}{2i\pi}\int_{\mathcal{C}(r)} c(\omega,1) g\Big(\frac{z}{\omega}\Big) d\omega, \quad |r| = 1,\ \forall |z| \leq 1, \tag{2.3.13}$$

qui se réécrit sous la forme

$$z\frac{\partial c(z,1)}{\partial z} = -c(z,1)\int_0^z c(\omega,1)\log\Big(\frac{z}{\omega}\Big)\, d\omega, \tag{2.3.14}$$

elle-même équivalente au système différentiel non linéaire

$$\begin{cases} z\dfrac{\partial c(z,1)}{\partial z} + c(z,1)v(z,1) = 0, \\ z\dfrac{\partial^2 v(z,1)}{\partial z^2} + \dfrac{\partial v(z,1)}{\partial z} = c(z,1), \end{cases} \tag{2.3.15}$$

sous les conditions initiales

$$v(0,1) = 0, \quad \frac{\partial v(z,1)}{\partial z}_{|z=0} = 1, \quad c(0,1) = 1.$$

(c) En outre, $c(z,1)$ est analytique dans le plan complexe ouvert, sauf en un point réel négatif q, et on a

$$c(z,1) \neq 0,\ \forall z \neq q \cup \infty.$$

Preuve. Les trois premières équations dans (2.3.12) découlent directement des relations définissant les $a_k(\rho)$ (cf. (2.3.7), (2.3.10)), la quatrième provenant de (2.3.9) par prolongement analytique. L'existence et l'unicité sont immédiates grâce aux relations de convolution (2.3.7).

Les propriétés ayant trait à la morphologie de $c(z,\rho)$, $\rho \leq 1$, sont délicates. D'abord le lecteur constatera que l'équation (2.3.13) s'obtient rigoureusement par continuité à partir de (2.3.7) (ou de (2.3.8), mais non de (2.3.12) !). Ensuite, il y un véritable *changement de phase* lorsque $\rho = 1$. Nous reviendrons sur cette question dans la section 2.3.3. ■

Il est intéressant de noter que la fonction $w(y) \stackrel{\text{def}}{=} v(e^y, 1) + 1$ satisfait l'équation non linéaire du troisième ordre, dite de *Blasius* [2],

$$w'''(y) + w(y)w''(y) = 0,$$

qui intervient en hydrodynamique, pour décrire des écoulements laminaires stationnaires avec couches limites ! Sa solution explicite est toujours inconnue, bien que de nombreux auteurs l'aient étudiée au cours du siècle dernier (voir à ce sujet Schlichting [13] et Steinheuer [14]).

Au vu du lemme 2.3.3, nous pouvons maintenant esquisser les lignes de forces de la méthode proposée. Le nœud Gordien qu'il faut trancher est l'évaluation de $\xi(\rho)$, défini par (2.3.11). À cette fin, on exploitera *l'équation d'ancrage*

$$c(\xi(\rho), \rho) = 0, \tag{2.3.16}$$

qui découle de la remarque faite à la fin de la section précédente et de la première équation de (2.3.12). D'après la forme de la troisième équation de (2.3.12), la plus petite solution positive de $c(u, \rho) = 0$ vérifie

$$\frac{-1}{u \log \rho} = (\log \rho)^2 \sum_{i \geq 1} i\rho^{i+1} c(u\rho^{i+1}, \rho),$$

dont le membre droit représente une fonction analytique, bornée dans tout compact $\forall \rho \leq 1$: lorsque $\rho \to 1$, alors nécessairement $u \to \infty$ et tous les zéros positifs de l'équation d'ancrage sont rejetés à l'infini. La clé du problème consiste donc à trouver le comportement asymptotique en z des diverses fonctions, dans un domaine conique $0 \leq z \leq \mathcal{U}(\rho)$, incluant $\xi(\rho)$: dans ce cône, $c(z, \rho)$ est, en un certain sens, très proche de $c(z, 1)$ et on a

$$v(z, \rho) \approx w(z, \rho) \stackrel{\text{def}}{=} \log^2 \rho \sum_{i \geq 1} i\rho^{i+1} c(z\rho^{i+1}, 1). \tag{2.3.17}$$

Or $w(z, 1)$ a un comportement logarithmique en z, si bien que $\xi(\rho)$ se déduit par inversion directe. Les lignes principale de la preuve sont données ci-dessous.

Nous utiliserons la transformée de Mellin (cf. par exemple [5]) de $c(z, 1)$, définie par

$$c^*(s) = \int_0^\infty x^{s-1} c(x) dx.$$

Le comportement de $c(z, 1)$ dans la région $\Re(z) > 0$, donné dans la section 2.3.3, permet d'affirmer l'existence de $c^*(s, 1), \forall s$, $\Re(s) > 0$, et même de tous les moments de $c(z, 1)$ sur l'axe réel positif.

Lemme 2.3.4. *La fonction $w(z, \rho)$ définie par (2.3.17) admet, $\forall z$, $\Re(z) > 0$, le développement*

$$\begin{aligned} w(z, \rho) = c^*(1, 1) \log(\rho z) - \frac{\partial c^*(1, 1)}{\partial s} \\ + (\log \rho)^2 \big[(\log z)\Phi(z, \rho) + \Psi(z, \rho)\big] + \mathcal{O}(z^{-d}), \end{aligned} \tag{2.3.18}$$

où d est un nombre positif arbitraire, Φ et Ψ sont, $\forall \rho \leq 1$, des fonctions fluctuantes, nulles pour $\rho = 1$ et dont les oscillations sont de faible amplitude sinon.

Preuve. D'après les formules d'inversion de la transformée de Mellin,

$$w(z,\rho) = \frac{1}{2i\pi}\int_{\sigma-i\infty}^{\sigma+i\infty} z^{-s}c^*(s+1,1)\Big[\frac{\log\rho}{1-\rho^s}\Big]^2 ds, \quad \forall\sigma\in(-1,0). \tag{2.3.19}$$

Soit

$$s_n \stackrel{\text{def}}{=} \frac{2in\pi}{\log\rho}, \quad \forall n\in\mathbb{Z}.$$

On applique le théorème de Cauchy à (2.3.19), en prenant comme nouveau contour d'intégration la droite verticale d'abscisse $\Re(s) = d > 0$. Il vient alors

$$w(z,\rho) = \sum_{n\in\mathbb{Z}} z^{-s_n}\Big[c^*(s_n+1,1)\log(\rho z) - \frac{\partial c^*(s_n+1,1)}{\partial s}\Big] + \mathcal{O}(z^{-d}). \tag{2.3.20}$$

La série dans (2.3.20) est égale à la somme des résidus pris sur la ligne verticale d'abscisse $\Re(s) = 0$ et elle est uniformément bornée, $\forall\rho\le 1$. En effet, en intégrant par parties, on a l'inégalité

$$|c^*(s_n+1,1)| = \Big|\int_0^\infty y^{s_n}c(y,1)dy\Big| \le \Big|\frac{\Gamma(s_n+1)}{\Gamma(s_n+k+1)}\Big|\int_0^\infty y^k|c^{(k)}(y,1)|dy,$$

où $c^{(k)}$ désigne la dérivée d'ordre k de $c(z,1)$ et $\Gamma(x)$ est la fonction Eulérienne habituelle. Or

$$\Big|\frac{\Gamma(s_n+1)}{\Gamma(s_n+k+1)}\Big| < \frac{\log^k\rho}{k!},$$

ce qui démontre (2.3.18) et le lemme. ■

Théorème 2.3.5. *On a, pour un réel $d>1$, les développements*

$$\rho\xi(\rho) = \exp\Big[-\frac{1}{\log\rho\, c^*(1,1)} - \frac{\partial\log c^*(1,1)}{\partial s}\Big]\big(1+\mathcal{O}((\log\rho)^d)\big), \tag{2.3.21}$$

$$\alpha'(1,\rho) = \frac{1}{(\log\rho)^2c^*(1,1)} + \mathcal{O}((\log\rho)^d). \tag{2.3.22}$$

Preuve. D'après (2.3.16), $\xi(\rho)$ est solution de l'équation en x

$$1+\log\rho\, v(x,\rho) = 0.$$

Une analyse poussée, dont la preuve n'est pas incluse ici, permet d'affirmer qu'il est loisible de remplacer cette équation par l'équation localement équivalente

$$1+\log\rho\, w(x,\rho) + \mathcal{O}\big((\log\rho)^p\big) = 0, \tag{2.3.23}$$

où p est un nombre positif, $p>1$. Il vient alors, d'après le lemme 2.3.4,

$$-\frac{1}{\log\rho} = \log(\rho x)c^*(1,1) - \frac{\partial c^*(1,1)}{\partial s} + \log(\rho x)\Phi(x,\rho) + \Psi(x,\rho) + \mathcal{O}(x^{-d}), \tag{2.3.24}$$

ce qui donne (2.3.21).

En ce qui concerne $\alpha'(1,\rho)$, on applique la relation simple

$$\rho\alpha'(1,\rho) = \frac{\alpha_1}{\alpha_0} - \alpha_1,$$

obtenue par simple dérivation de (2.3.2) en $z=0$, et qui conduit à (2.3.22). ■

2.3.3 Remarques et compléments

Il est important de noter que l'une des difficultés techniques majeures du problème, au delà de son caractère fortement non linéaire, provient du *changement de phase* qui surgit lorsque $\rho = 1$. De fait, la fonction $c(z,\rho)$ reste positive jusqu'au point $\xi(\rho)$, où elle s'annule ; ensuite ses amplitudes restent faibles, puis, brusquement, pour des valeurs de z très supérieures à $\exp\big(\xi(\rho)\big)$, elle se met à osciller fortement, avec des amplitudes non bornées, ce qui implique en particulier que $\int_0^\infty c(x,\rho)dx$ n'existe pas.

Par contre, lorsque $\rho = 1$, l'intégrale $\int_0^\infty c(x,1)dx = A$ est finie, il n'y a plus de phénomènes oscillatoires notables, mais la fonction $c(x,1)$ n'est plus entière : elle a une singularité (qui semble être un pôle d'ordre trois) située sur l'axe réel négatif et elle ne s'annule pas pour $z \geq 0$. On peut montrer, dans le demi-plan $\Re(z) > 0$, les comportements suivants :

$$\begin{cases} c(z,1) = \exp\left[-\dfrac{c^*(1,1)}{2}\log^2 z + B\log z + \dfrac{D\log z}{z^2} + o\Big(\dfrac{\log z}{z^2}\Big)\right], \\ v(z,1) = c^*(1,1)\log z + B + \dfrac{D\log z}{z^2} + o\left(\dfrac{\log z}{z^3}\right), \end{cases}$$

où B et D sont des constantes.

Enfin, il serait possible d'améliorer certaines des estimations données dans la section précédente en mettant $c(z,\rho)$ sous la forme (par itération de (2.3.12))

$$c(z,\rho) = c(z\rho^{I+1},\rho)\prod_{i=0}^{I}\big[1 + \log\rho\, v(z\rho^i,\rho)\big],$$

où I représente un entier positif quelconque. Le produit ci-dessus est uniformément convergent, $\forall I \leq \infty$, pour tout z dans un compact du plan complexe, car il se comporte comme la série

$$z(\log\rho)^3\sum_{k\geq 0}\sum_{i\geq 1} i\rho^{i+k+1}c(z\rho^{i+k+1},\rho),$$

qui, d'après le principe du module maximum pour les fonctions analytiques, converge uniformément, son module étant majoré par $|z\rho^2 c(\rho z,\rho)|$. L'idée consisterait à choisir I tel que

$$0 < z\rho^I \leq 1, \quad \text{i.e.} \quad I \leq -\frac{\log z}{\log\rho},$$

puis d'utiliser les propriétés de la fonction Eulérienne Γ.

2.4 Un réseau étoilé partiellement symétrique[4]

Dans la section 2.1, nous avons uniquement considéré le problème de partage de la bande passante pour des réseaux symétriques. Cependant, on peut aussi envisager la structure en étoile comme un possible modèle de communication et de transfert sur Internet : les liens « entrants »

4. Auteurs : G. Fayolle et J.-M. Lasgouttes

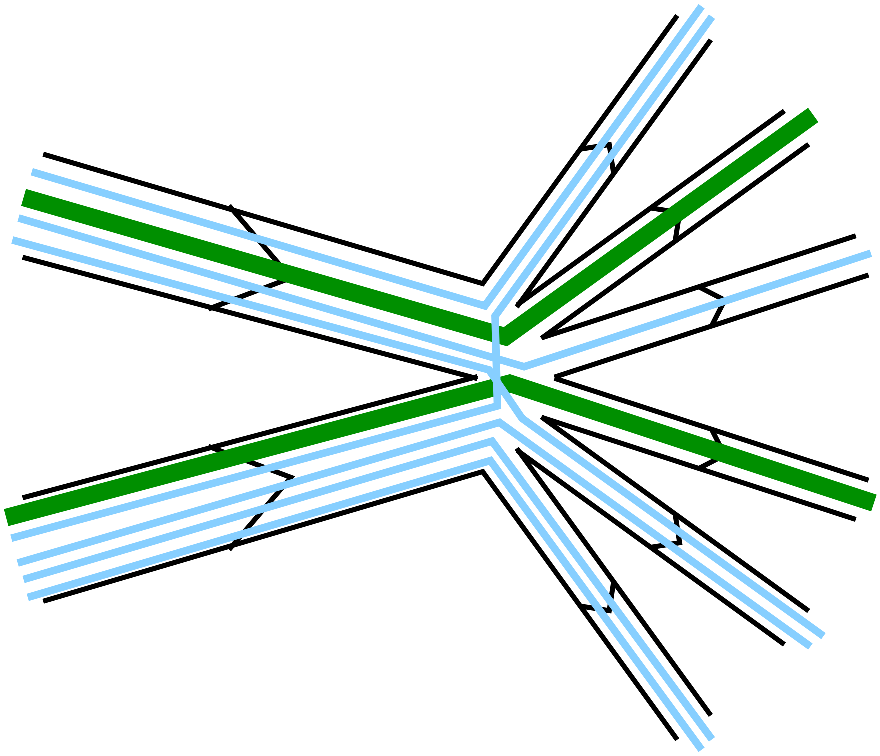

FIG. 2.4.1: *Le réseau en étoile asymétrique.*

correspondent alors aux processeurs de serveurs WEB, tandis que les liens « sortants » représentent les arrivées vers les clients destinataires. Dans ce cas, il faut abandonner les hypothèses de symétrie entre les deux familles de liens, chacune d'elle pouvant provoquer des saturations de nature différente.

On étudie le réseau représenté sur la figure 2.4.1, qui comporte N^{in} liens entrants et N^{out} liens sortants, de capacités respectives C^{in} et C^{out}. Une route est formée d'un couple ordonné {lien entrant, lien sortant}

On introduit les quantités suivantes :

- l'intensité des arrivées sur chaque route $\lambda_r \stackrel{\text{def}}{=} \dfrac{\lambda}{N^{\text{in}}}$;
- σ, longueur moyenne d'un message ;
- $\varepsilon \stackrel{\text{def}}{=} \dfrac{C^{\text{out}}}{C^{\text{in}}}$.

On suppose que λ et σ sont identiques pour chaque route, que la capacité C^{out} est donnée et que la charge $\rho^{\text{out}} \stackrel{\text{def}}{=} \lambda\sigma/C^{\text{out}}$ est plus petite que 1.

Ensuite, la capacité C^{in} est choisie de façon à garantir que la charge *entrante*

$$\rho^{\text{in}} \stackrel{\text{def}}{=} \frac{\lambda\sigma N^{\text{out}}}{C^{\text{in}} N^{\text{in}}},$$

ait une limite bornée strictement inférieure à 1 quand les nombres N^{in} et N^{out} augmentent indéfiniment (le système est donc ergodique). Lorsque le rapport $N^{\text{in}}/N^{\text{out}}$ tend vers zéro, les liens entrants (qui correspondent par exemple à un *backbone*) ont beaucoup de connexions actives et une grande capacité, alors que les liens sortants sont eux en régime normal.

Par une approche calquée sur celle de la section 2.1, il est possible d'obtenir le jeu d'équations heuristiques suivant :

$$\begin{cases} \alpha_{k+1} v_{k+1} = \rho^{\text{in}} \bar{\beta} \alpha_k, \\ \beta_{k+1} u_{k+1} = \rho^{\text{in}} \bar{\beta} \beta_k, \end{cases} \tag{2.4.1}$$

où

$$\begin{cases} u_k = \sum_{\ell \geq 0} \alpha_\ell \min[k, \varepsilon \ell], \\ v_k = \sum_{\ell \geq 0} \beta_\ell \min[\varepsilon k, \ell], \end{cases} \tag{2.4.2}$$

et α_k (resp. β_k) représente la proportion de liens entrants (resp. sortants) avec k connexions actives.

Par la suite, on va tenter d'étudier le comportement du système lorsque $\varepsilon \to 0$, ce qui, dans une certaine mesure, traduira l'influence respective des parties « in » et « out » lorsque la taille du réseau augmente.

En introduisant les fonctions génératrices (avec des notations naturelles où, sauf mention explicite, le nom de la fonction correspond à celui des coefficients, i.e. $f(z) = \sum_{k \geq 0} f_k z^k$), on obtient le système intégro-fonctionnel

$$\begin{cases} z\rho^{\text{in}} \bar{\beta} \alpha(z) = \dfrac{1}{2i\pi} \displaystyle\int_{\mathcal{C}(r)} \alpha\Big(\frac{z}{\omega}\Big) \frac{v(\omega)}{\omega} d\omega \,, \\ z\rho^{\text{in}} \bar{\beta} \beta(z) = \dfrac{1}{2i\pi} \displaystyle\int_{\mathcal{C}(r)} \beta\Big(\frac{z}{\omega}\Big) \frac{u(\omega)}{\omega} d\omega \,, \end{cases} \tag{2.4.3}$$

où l'intégration est prise sur un cercle de rayon $r < 1$ centré à l'origine. Au vu des études précédentes, la résolution exacte du système (2.4.4) est humainement hors de portée !

Formellement, le problème d'écriture de $u(z)$ et $v(z)$ est lié à l'expression de sommes de Dirichlet du genre

$$A(x) = \sum_{\lambda_n \leq x} a_n.$$

La formule suivante, dite de Perron—dans laquelle $d(s) = \sum_{n \geq 1} a_n e^{-\lambda_n s}$, x est un réel positif et les λ_n forment une suite de nombres positifs croissant vers l'infini (cf. [16])—répond précisément à la question :

$$\frac{A(x+0) + A(x-0)}{2} = \frac{1}{2i\pi} \lim_{T \to \infty} \int_{\sigma - iT}^{\sigma + iT} \frac{d(s)}{s} e^{sx} ds \,, \tag{2.4.4}$$

σ étant une abscisse de convergence (réelle positive) pour la série de Dirichlet définissant la fonction $d(s)$.

Pour toute fonction analytique $f(z) = \sum f_n z^n$, on notera $\hat{f}(s) \stackrel{\text{def}}{=} f(e^{-s})$ et $\hat{f}'(s)$ la dérivée de $\hat{f}(s)$ par rapport à s. Il vient alors, en appliquant (2.4.4) dans la région $|e^{\sigma/\varepsilon}\omega| < 1$ à la suite $\lambda_n \equiv n$ et en supposant ε *irrationnel*,

$$\begin{aligned} u(\omega) &= \sum_{k\geq 0} \omega^k \Big[\varepsilon \sum_{\ell\varepsilon\leq k} \ell\alpha_\ell + k \sum_{\ell\varepsilon>k} \alpha_\ell\Big] \\ &= -\frac{1}{2i\pi}\int_{\sigma-i\infty}^{\sigma+i\infty}\left[\frac{\varepsilon\hat{\alpha}'(s)}{s(1-\omega e^{s/\varepsilon})} + \frac{\omega e^{s/\varepsilon}\hat{\alpha}(s)}{s(1-\omega e^{s/\varepsilon})^2}\right]ds + \frac{\omega\alpha(1)}{(1-\omega)^2} \\ &= -\frac{1}{2i\pi}\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{\varepsilon\hat{\alpha}(s)ds}{s^2(1-\omega e^{s/\varepsilon})} + \frac{\omega\alpha(1)}{(1-\omega)^2}, \end{aligned} \tag{2.4.5}$$

l'équation (2.4.5) s'obtenant à l'aide d'une intégration par parties. De manière analogue, on a

$$\begin{aligned} v(\omega) &= \sum_{k\geq 0} \omega^k \Big[\sum_{\ell\leq k\varepsilon} \ell\beta_\ell + \varepsilon k \sum_{\ell>k\varepsilon} \beta_\ell\Big] \\ &= -\frac{1}{2i\pi}\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{\hat{\beta}(s)ds}{s^2(1-\omega e^{s\varepsilon})} + \frac{\varepsilon\omega\beta(1)}{(1-\omega)^2}. \end{aligned} \tag{2.4.6}$$

On porte les équations (2.4.5) et (2.4.6) dans le système (2.4.3), on permute ad libitum l'ordre des intégrations (dans les régions autorisées), puis on extrait les résidus au point d'affixe 1 [lesquels proviennent des fractions en $(1-\omega)^{-2}$ dans les membres droits de (2.4.5) et (2.4.6)] pour finalement obtenir

$$\rho^{\text{in}}\bar{\beta}z\alpha(z) = \varepsilon\beta(1)z\alpha'(z) - \frac{\varepsilon}{2i\pi}\int_{\varepsilon\sigma_1-i\infty}^{\varepsilon\sigma_1+i\infty}\frac{\alpha(ze^t)\,\beta(e^{-t/\varepsilon})\,dt}{t^2} \tag{2.4.7}$$

$$\rho^{\text{in}}\bar{\beta}z\beta(z) = \alpha(1)z\beta'(z) - \frac{1}{2i\pi}\int_{\frac{\sigma_2}{\varepsilon}-i\infty}^{\frac{\sigma_2}{\varepsilon}+i\infty}\frac{\beta(ze^t)\,\alpha(e^{-t\varepsilon})\,dt}{t^2}. \tag{2.4.8}$$

Dans le système ci-dessus, les nombres σ_1 et σ_2 sont positifs et satisfont les inégalités

$$|z|\,e^{\sigma_1\varepsilon} \leq 1, \quad |z|\,e^{\frac{\sigma_2}{\varepsilon}} \leq 1.$$

La dépendance en ε des fonctions β et α est très complexe, mais le comportement limite lorsque $\varepsilon \to 0$ peut se déduire de (2.4.7) et (2.4.8). Désormais, nous ferons systématiquement apparaître ε comme une variable explicite dans les fonctions et quantités concernées (par ex. $\alpha(z,\varepsilon), \bar{\beta}(\varepsilon)$, etc.).

À partir des systèmes (2.4.1) et (2.4.2), et en utilisant la relation de conservation

$$\rho^{\text{out}}\varepsilon\alpha'(1,\varepsilon) = \rho^{\text{in}}\beta'(1,\varepsilon),$$

il est facile de voir que

$$\lim_{k\to\infty}\frac{\alpha_{k+1}(\varepsilon)}{\alpha_k(\varepsilon)} = \rho^{\text{in}}, \quad \lim_{k\to\infty}\frac{\beta_{k+1}(\varepsilon)}{\beta_k(\varepsilon)} = \rho^{\text{out}}, \quad \forall\varepsilon > 0, \tag{2.4.9}$$

les convergences étant évidemment *non uniformes* en ε.

Dans la région d'ergodicité $\rho^{\text{in}} < 1$, $\rho^{\text{out}} < 1$, on a $\alpha(1,\varepsilon) = \beta(1,\varepsilon) = 1$ et les fonctions $\alpha(z,\varepsilon)$ et $\beta(z,\varepsilon)$ ont, d'après (2.4.9), des rayons de convergence respectifs au moins égaux à $1/\rho^{\text{in}}$ et $1/\rho^{\text{out}}$. Il devient alors possible, dans (2.4.7) et (2.4.8), de décaler les droites verticales d'intégration légèrement à gauche de l'origine, ce qui conduit au système

$$\begin{cases} \bar{\beta}(\varepsilon)(1 - z\rho^{\text{in}})\alpha(z,\varepsilon) = \dfrac{\varepsilon}{2i\pi}\displaystyle\int_{\varepsilon\delta_1 - i\infty}^{\varepsilon\delta_1 + i\infty} \frac{\alpha(ze^t,\varepsilon)\,\beta(e^{-t/\varepsilon},\varepsilon)\,dt}{t^2}, \\ \dfrac{\rho^{\text{in}}\bar{\beta}(\varepsilon)}{\rho^{\text{out}}}(1 - z\rho^{\text{out}})\beta(z,\varepsilon) = \dfrac{1}{2i\pi}\displaystyle\int_{\delta_2 - i\infty}^{\delta_2 + i\infty} \frac{\beta(ze^t,\varepsilon)\,\alpha(e^{-t\varepsilon},\varepsilon)\,dt}{t^2}. \end{cases} \tag{2.4.10}$$

δ_1 et δ_2 étant réels négatifs, de modules suffisamment faibles pour que les lignes d'intégration soient dans les domaines d'analycité des fonctions $\alpha(z,\varepsilon)$ et $\beta(z,\varepsilon)$. En outre, grâce à (2.4.10), on pourrait effectuer le prolongement analytique des différentes fonctions. Disons simplement qu'elles sont méromorphes, leurs pôles étant nécessairement de la forme $(\rho^{\text{in}})^m, (\rho^{\text{out}})^n$, où m et n sont des entiers positifs.

On souhaite trouver les distributions limites, lorsque $\varepsilon = 0$. Nous énoncerons quelques conjectures, en esquissant uniquement les lignes de force du raisonnement mathématique. D'abord, la fonction $\alpha(z,\varepsilon)$ dégénère (à l'instar de la variable aléatoire $X_\ell(\varepsilon)$) : il y a un changement de phase très marqué et tous les pôles à distance finie disparaissent.

Par contre, la fonction $\beta(z,0) = \lim_{\varepsilon\to 0}\beta(z,\varepsilon)$ correspond à une distribution propre et n'a qu'un seul pôle à distance finie au point $x = 1/\rho^{\text{out}}$.

On effectue dans la première équation de (2.4.10) le changement de variable $z \to z^\varepsilon$ et on introduit la fonction $\gamma(z,\varepsilon) \stackrel{\text{def}}{=} \alpha(z^\varepsilon,\varepsilon)$. Puisque la fonction z^ε a un nombre infini de déterminations (elle n'est pas analytique dans un voisinage de l'origine), on suppose que z reste dans un demi-plan fixé. Il vient alors

$$\bar{\beta}(\varepsilon)(1 - z^\varepsilon\rho^{\text{in}})\gamma(z,\varepsilon) = \frac{1}{2i\pi}\int_{\delta_1 - i\infty}^{\delta_1 + i\infty} \frac{\gamma(ze^t,\varepsilon)\,\beta(e^{-t},\varepsilon)\,dt}{t^2}, \tag{2.4.11}$$

$$\frac{\rho^{\text{in}}\bar{\beta}(\varepsilon)}{\rho^{\text{out}}}(1 - z\rho^{\text{out}})\beta(z,\varepsilon) = \frac{1}{2i\pi}\int_{\delta_2 - i\infty}^{\delta_2 + i\infty} \frac{\beta(ze^t,\varepsilon)\,\gamma(e^{-t},\varepsilon)\,dt}{t^2}. \tag{2.4.12}$$

Pour z fixé, il *semblerait* que l'on puisse obtenir, à l'aide d'un prolongement analytique,

$$\begin{cases} \lim\limits_{\varepsilon\to 0}\gamma(z,\epsilon) = z^u, \\ \lim\limits_{\varepsilon\to 0}\beta(z,\epsilon) = \dfrac{Q_v(z)}{1 - z\rho^{\text{out}}}, \end{cases} \tag{2.4.13}$$

où $u = \rho^{\text{in}}\bar{\beta}/\rho^{\text{out}}$ et $Q_v(z)$ est un polynôme de degré v, se réduisant possiblement à une constante, auquel cas on retrouve le phénomène d'insensibilité de la section 2.6.2. Cette analyse sera poursuivie ultérieurement.

2.5 Champ moyen de second ordre pour un grand réseau symétrique[5]

Même si les comparaisons entre les résultats de simulation de la section 1.2 et les calculs numériques semblent montrer une convergence quand $N \to \infty$, il est intéressant de voir ce qui se passe quand on ne suppose plus l'indépendance conditionnelle des deux liens traversés par une connexion active. En fait, la section 2.5.2 montre que les équations obtenues en section 2.1 (et dérivées d'une manière différente en section 2.5.1) ne sont pas exactes, et qu'il est possible de les améliorer en considérant un champ moyen d'ordre 2. Nous verrons toutefois en section 2.5.3 que le champ moyen considéré précédemment est une assez bonne approximation, surtout si l'on considère la complexité ajoutée par le second modèle.

Le modèle et les notations sont les mêmes qu'en section 2.1.1, à ceci près que l'on se restreint ici au cas $L = 2$ et que l'on note $\zeta(X^{(N)}_{r(1)}, X^{(N)}_{r(2)})$ la bande passante allouée à une route $r \in \mathcal{R}$, que l'on supposera uniquement dépendante de l'état des liens visités par cette route. Dans le cas de la politique min, on a

$$\zeta(x, y) = \frac{1}{x} \wedge \frac{1}{y}.$$

2.5.1 Équations de récurrence – ordre 1

On définit le *champ moyen* $\alpha^{(N)}_x$ comme au début de la section 2.1.2. Par contre, pour calculer l'intensité de départ d'un lien, nous introduisons mesure suivante, qui représente la proportion de connexions actives dont les extrémités ont respectivement x et y connexions actives :

$$\beta^{(N)}_{xy} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i,j \in \mathcal{L}} x_{ij} \mathbb{1}_{\{X^{(N)}_i = x, X^{(N)}_j = y\}},$$

où x_{ij} vaut $x^{(N)}_r$ si la route r passe par i puis j, 0 sinon. Il est intéressant de noter que

$$\sum_{y>0} \beta^{(N)}_{xy} = x\alpha^{(N)}_x. \tag{2.5.1}$$

Si on note $G^{(N)}$ le générateur de la chaîne de Markov associée au système, alors son application à $\alpha^{(N)}_x$ donne pour $x \geq 0$

$$G^{(N)}[\alpha^{(N)}_x] = \frac{\lambda}{N}\left[\alpha^{(N)}_{x-1} - \alpha^{(N)}_x\right] + \frac{1}{N} \sum_{y>0} \left[\beta^{(N)}_{x+1,y} \frac{1}{\sigma} \zeta(x+1, y) - \beta^{(N)}_{xy} \frac{1}{\sigma} \zeta(x, y)\right].$$

Par sommation, on obtient

$$\rho \mathbb{E}\alpha^{(N)}_x = \sum_{y>0} \mathbb{E}\beta^{(N)}_{x+1,y} \zeta(x+1, y),$$

5. Auteurs : F. Delcoigne et A. de La Fortelle

et, en supposant que $\beta_{xy}^{(N)}$ converge quand $N \to \infty$ vers $\beta_{xy} \in \mathbb{R}$,

$$\rho\alpha_x = \sum_{y>0} \beta_{x+1,y}\zeta(x+1,y). \tag{2.5.2}$$

Pour pouvoir rendre cette équation fermée en α_x, il est nécessaire de faire une hypothèse d'indépendance que nous appellerons d'ordre 1 :

Hypothèse H1. *β_{xy} a une forme produit en x et y. D'après (2.5.1), cette forme est nécessairement*

$$\beta_{xy} = \frac{x\alpha_x y\alpha_y}{\bar{\alpha}}. \tag{2.5.3}$$

Une manière de comprendre cette hypothèse est de remarquer que, quand $N \to \infty$, chaque route est pratiquement vide et $x_r^{(N)} \approx \mathbb{1}_{\{x_r^{(N)}>0\}}$. Alors on a,

$$\beta_{xy} = \mathbb{P}\big(X_{r(1)} = x, X_{r(2)} = y \;\big|\; x_r > 0\big), \tag{2.5.4}$$

et H1 revient à supposer que les lois des liens traversés par une connexion active sont indépendantes. Cette hypothèse est bien sûr fausse en général, mais on peut se persuader qu'elle devient d'autant plus raisonnable que ρ est proche de 0 ou de 1.

Sous H1, la relation (2.5.2) devient

$$\rho\bar{\alpha}\alpha_x = \alpha_{x+1} \sum_{y>0} (x+1)y\alpha_y\zeta(x+1,y). \tag{2.5.5}$$

2.5.2 Équations de récurrence – ordre 2

Afin de vérifier la pertinence de H1, nous cherchons maintenant à écrire l'équation d'évolution de β_{xy}, pour vérifier si (2.5.3) en est solution.

Soit $r = (i,j)$ une route quelconque et x et y des entiers fixés et soit $\delta_r(x,y) \stackrel{\text{def}}{=} \mathbb{1}_{\{X_{r(1)}^{(N)}=x, X_{r(2)}^{(N)}=y\}}$. On a alors

$$\begin{aligned}
&G^{(N)}[x_r^{(N)}\delta_r(x,y)] \\
&\quad = \frac{\lambda}{R^{(N)}}\Big[(x_r^{(N)}+1)\delta_r(x-1,y-1) - x_r^{(N)}\delta_r(x,y)\Big] \\
&\qquad + \frac{\lambda(R^{(N)}-1)}{R^{(N)}} x_r^{(N)}\big[\delta_r(x-1,y) - \delta_r(x,y)\big] \\
&\qquad + \frac{\lambda(R^{(N)}-1)}{R^{(N)}} x_r^{(N)}\big[\delta_r(x,y-1) - \delta_r(x,y)\big] \\
&\qquad + \frac{x_r^{(N)}}{\sigma}\Big[\zeta(x+1,y+1)(x_r^{(N)}-1)\delta_r(x+1,y+1) - \zeta(x,y)x_r^{(N)}\delta_r(x,y)\Big] \\
&\qquad + \frac{x_r^{(N)}}{\sigma} \sum_{r'\ni i, r'\neq r} x_{r'}^{(N)}\Big[\zeta(x+1,X_{r'(2)}^{(N)})\delta_r(x+1,y) - \zeta(x,X_{r'(2)}^{(N)})\delta_r(x,y)\Big] \\
&\qquad + \frac{x_r^{(N)}}{\sigma} \sum_{r'\ni j, r'\neq r} x_{r'}^{(N)}\Big[\zeta(x+1,X_{r'(1)}^{(N)})\delta_r(x+1,y) - \zeta(x,X_{r'(1)}^{(N)})\delta_r(x,y)\Big].
\end{aligned}$$

De même que l'équation d'évolution de $\alpha_x^{(N)}$ faisait intervenir la mesure empirique $\beta_{xy}^{(N)}$, nous avons maintenant besoin de la mesure

$$\gamma_{x,y,z}^{(N)} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i,j\in\mathcal{L}} \sum_{k\neq j} x_{ij}^{(N)} x_{ik}^{(N)} \mathbb{1}_{\{X_i^{(N)}=x, X_j^{(N)}=y, X_k^{(N)}=z\}}.$$

On supposera, de même que pour les α et β, que cette mesure a une limite déterministe $\gamma_{x,y,z}$ quand $N \to \infty$. De plus, nous faisons l'hypothèse que, asymptotiquement, $x_r^{(N)}$ vaut soit 0 soit 1, si bien que $x_r^{(N)}(x_r^{(N)}-1)=0$ (ce qui est plausible considérant que le nombre total de routes est bien plus grand que le nombre total de liens).

Alors

$$\begin{aligned}\lim_{N\to\infty} G^{(N)}\beta_{xy}^{(N)} &= \lambda\alpha_{x-1}\alpha_{y-1}\\ &\quad + \lambda[\beta_{x-1,y} - \beta_{xy} + \beta_{x,y-1} - \beta_{xy}]\\ &\quad - \frac{1}{\sigma}\zeta(x,y)\beta_{xy}\\ &\quad + \frac{1}{\sigma}\sum_{z>0}\Big[\gamma_{x+1,y,z}\zeta(x+1,z) - \gamma_{x,y,z}\zeta(x,z)\\ &\qquad\qquad + \gamma_{y+1,x,z}\zeta(y+1,z) - \gamma_{y,x,z}\zeta(y,z)\Big],\end{aligned}$$

et donc

$$\begin{aligned}0 &= \rho\Big[\alpha_{x-1}\alpha_{y-1} + \beta_{x-1,y} - \beta_{xy} + \beta_{x,y-1} - \beta_{xy}\Big] - \zeta(x,y)\beta_{xy}\\ &\quad + \sum_{z>0}\Big[\gamma_{x+1,y,z}\zeta(x+1,z) - \gamma_{x,y,z}\zeta(x,z)\Big]\\ &\quad + \sum_{z>0}\Big[\gamma_{y+1,x,z}\zeta(y+1,z) - \gamma_{y,x,z}\zeta(y,z)\Big].\end{aligned} \tag{2.5.6}$$

Pour rendre l'équation (2.5.6) fermée, il faut encore une fois recourir à une hypothèse d'indépendance, que nous appellerons d'ordre 2. Considérons d'abord les propriétés intrinsèques de $\gamma_{x,y,z}$: similairement à (2.5.1), nous avons des relations entre α, β et γ :

$$\sum_{z>0} \gamma_{x,y,z}^{(N)} = (x-1)\beta_{xy}^{(N)}, \tag{2.5.7}$$

$$\sum_{y>0,z>0} \gamma_{x,y,z}^{(N)} = x(x-1)\alpha_x^{(N)}. \tag{2.5.8}$$

Notons que $\alpha^{(N)}$ peut être dérivée de $\beta^{(N)}$ par (2.5.1), et donc que $\alpha^{(N)}$ n'est pas une variable externe dans les équations. Il existe de plus une symétrie en y, z : $\gamma_{x,y,z} = \gamma_{x,z,y}$.

Hypothèse H2. *$\gamma_{x,y,z}$ a une forme Markovienne, c'est-à-dire que conditionnellement à x, y et z sont indépendants. Grâce à la symétrie et aux équations (2.5.7)–(2.5.8), on en déduit que*

$$\gamma_{x,y,z} = (x-1)\frac{\beta_{xy}\beta_{xz}}{x\alpha_x}, \quad \forall x \geq 2,\ \forall y,z \geq 1.$$

Définissons maintenant le taux moyen de départ

$$\hat{\zeta}(x) \stackrel{\text{def}}{=} (x-1) \sum_{y>0} \zeta(x,y) \frac{\beta_{xy}}{x\alpha_x}, \qquad \forall x \geq 1.$$

L'équation (2.5.6) devient, sous H2,

$$\begin{aligned} 0 \;=\; & \rho\Big[\alpha_{x-1}\alpha_{y-1} + \beta_{x-1,y} - \beta_{xy} + \beta_{x,y-1} - \beta_{xy}\Big] - \zeta(x,y)\beta_{xy} \\ & + \beta_{x+1,y}\hat{\zeta}(x+1) - \beta_{xy}\hat{\zeta}(x) + \beta_{y+1,x}\hat{\zeta}(y+1) - \beta_{yx}\hat{\zeta}(y). \end{aligned} \tag{2.5.9}$$

Pour tester la validité de H1, il suffit de remarquer que l'indépendance est un cas particulier de dépendance Markovienne. Cependant, il faudrait rigoureusement formuler H1 dans le cadre de la mesure $\gamma_{x,y,z}$, car l'indépendance des variables x, y, z deux à deux n'implique pas l'indépendance des trois variables. Nous nous plaçons donc sous une hypothèse à peine plus forte H1$'$, selon laquelle $\gamma_{x,y,z}$ a une forme produit. Dans ce cas, β_{xy} conserve la forme (2.5.3), et d'après (2.5.7)–(2.5.8),

$$\gamma_{x,y,z} = x(x-1)\alpha_x \frac{y\alpha_y}{\bar{\alpha}} \frac{z\alpha_z}{\bar{\alpha}}.$$

Notons que H1$'$ implique (2.5.5), qui se récrit sous la forme

$$\rho\beta_{xy} = \beta_{x+1,y}\hat{\zeta}(x+1,y).$$

Alors (2.5.9) implique

$$\rho\alpha_{x-1}\alpha_{y-1} = \frac{x\alpha_x y\alpha_y}{\bar{\alpha}}\zeta(x,y), \quad \forall x,y \geq 1.$$

Pour α déterminée par (2.5.5), cette relation n'est satisfaite que dans des conditions très particulières, par exemple si ζ a une forme produit. H1$'$ n'est donc pas valide pour la politique min.

La question qui se pose alors est de savoir si les équations d'ordre 2 (2.5.9) sont asymptotiquement exactes. Pour avoir un début de réponse à cette question, il faut écrire l'équation d'évolution de $\gamma_{x,y,z}$. Les équations (et les calculs intermédiaires) sont comme on pouvait s'y attendre très compliqués, et nous ne les reproduirons pas ici. La conclusion est toutefois que l'on obtient à l'ordre 3 un système différent, a priori plus précis que le champ moyen d'ordre 2. Il est en fait très probable qu'il ne soit pas possible d'écrire des équations de champ moyen d'ordre fini qui soient exactes. On peut alors se contenter des équations d'ordre 1, dont on verra à la section précédente qu'elles sont suffisamment précises en tant qu'heuristique.

2.5.3 Validation numérique

Les résultats obtenus dans les études précédentes semblant montrer que les équations de champ moyen d'ordre 1 sont une très bonne approximation des résultats de simulations, cette section vise à quantifier plus précisément les écarts obtenus, et à les comparer au champ moyen d'ordre 2.

À cet effet, définissons la distance en variation totale entre deux distributions $(\alpha_x, x \geq 0)$ et $(\alpha'_x, x \geq 0)$

$$d(\alpha, \alpha') \stackrel{\text{def}}{=} \frac{1}{2} \sum_{x \geq 0} |\alpha_x - \alpha'_x|.$$

La figure 2.5.1 montre les distances respectives entre les résultats de simulation, et les solutions de champ moyen d'ordre 1 et 2, pour différentes valeurs de ρ. La première remarque est que les équations d'ordre 2 donnent une distance à la simulation bien plus petite que celles d'ordre 1, quelque soit la valeur de ρ. En particulier, la figure 2.5.2, qui montre la différence entre les distributions des deux champs moyen et de la simulation (pour la valeur $\rho = 0{,}7$ qui maximise cette différence), montre que cette réduction est uniforme pour toute la distribution. Un autre enseignement intéressant est que, si le champ moyen d'ordre 1 semble être un minorant de la simulation, les équations d'ordre 2 paraissent, elles, être majorantes. Ces conclusions se retrouvent pour les autres valeurs de ρ.

Malgré ces différences de précision, les distances obtenues sous H1 (moins de 0,02) sont très faibles, et cette distance diminue pour les valeurs extremales de ρ. Une explication intuitive de ce phénomène est la suivante :

- quand $\rho \to 0$, chaque branche tend à avoir soit 0 soit 1 connexion ; la loi conditionnelle (2.5.4) est donc déterministe et vérifie H1 ;
- inversement, quand $\rho \to 1$, deux branches ayant une connexion en commun ont beaucoup d'autres connexions actives, ce qui dilue leur interaction.

Il est donc vraisemblable que les équations asymptotiques obtenues précédemment pour $\rho \to 1$ restent exactes sous l'hypothèse H2.

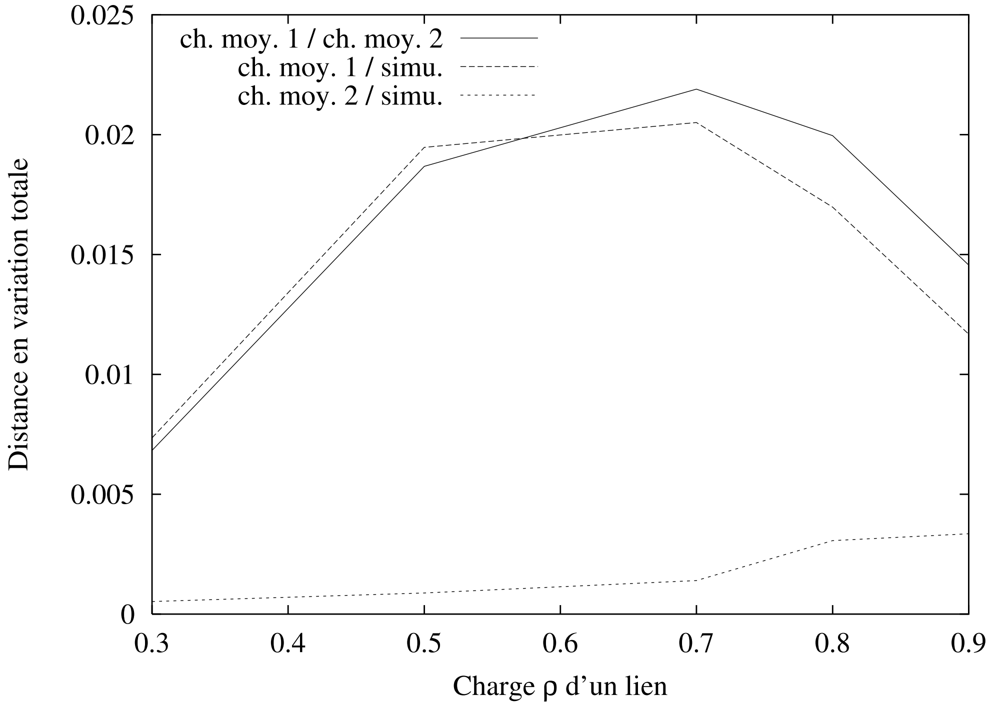

FIG. 2.5.1: *Distance en variation totale entre les lois α de la simulation ($N = 2000$) et des champs moyens d'ordre 1 et 2 , pour différentes valeurs de ρ.*

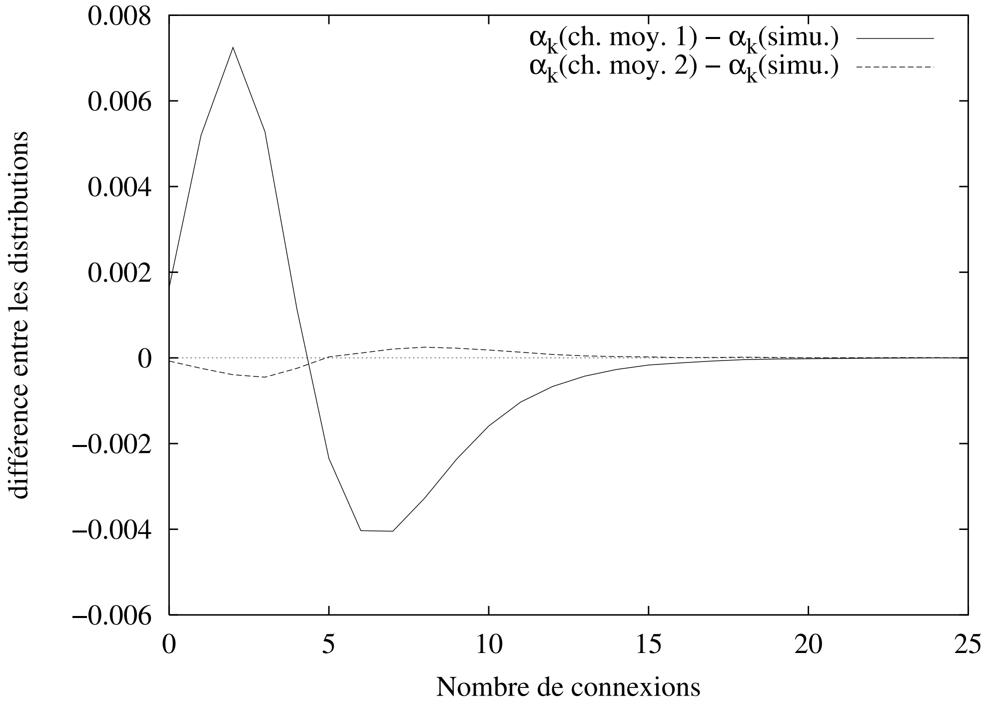

FIG. 2.5.2: *Différence entre les lois α du champ moyen d'ordre 1 (resp. d'ordre 2) et de la simulation ($N = 2000$), pour $\rho = 0{,}70$.*

2.6 Limite thermodynamique pour des systèmes markoviens non symétriques[6]

2.6.1 Présentation de la méthode

Nous ébauchons dans cette section une nouvelle approche de preuve de la propagation du chaos, d'abord d'un point de vue général, puis plus précisément appliquée aux cas particuliers d'un mécanisme de partage de bande passante (section 2.6.2) et d'un réseau de Jackson (section 2.6.3).

On considère un réseau $\mathcal{X}$ (de type file d'attente, au sens large) comportant un ensemble $\mathcal{N}$ de nœuds, avec $N \stackrel{\text{def}}{=} |\mathcal{N}|$. L'état du nœud $i \in \mathcal{N}$ à l'instant t étant noté $X_i(t) \in E$, on définit le processus

$$X(t) \stackrel{\text{def}}{=} (X_i(t))_{i\in\mathcal{N}}$$

qu'on suppose être une chaîne de Markov ergodique, de mesure invariante unique μ. Si nécessaire, on utilisera la notation $X^{(\mathcal{N})}(t)$ pour souligner la dépendance en $\mathcal{N}$.

Soit $\mathcal{B}(E)$ l'ensemble des fonctions réelles bornées pour la norme $\|\cdot\|_\infty$ définie par

$$\|f\|_\infty \stackrel{\text{def}}{=} \sup_{x\in E} f(x).$$

Si G est le générateur infinitésimal associé à $\mathcal{X}$, alors

$$\int_{E^N} Gf(x)d\mu(x) = 0,\ \forall f \in \mathcal{B}(E^N). \tag{2.6.1}$$

On notera aussi par la suite $\mathbb{E}^\mu$ l'espérance sous la mesure μ, de sorte que (2.6.1) peut être récrite

$$\mathbb{E}^\mu Gf(X) = 0,\ \forall f \in \mathcal{B}(E^N).$$

On veut étudier le comportement asymptotique de μ quand le volume de $\mathcal{X}$ augmente (i.e. quand le nombre de nœuds N devient grand). La méthode proposée est la suivante : on cherche d'abord à deviner quel est le comportement d'un nœud isolé au milieu du système, en supposant que ses interactions avec les autres nœuds sont indépendantes les unes des autres et suffisamment « diluées ». Soit $\tilde{G}_i$ le générateur associé à un tel nœud i. Sous des conditions d'ergodicité adéquates, il existe une mesure $\tilde{\mu}_i$ telle que

$$\mathbb{E}^{\tilde{\mu}_i}\tilde{G}_i f(X_i) = 0,\ \forall i \in \mathcal{N},\ \forall f \in \mathcal{B}(E). \tag{2.6.2}$$

À partir des $\tilde{G}_i$, on construit le générateur et la mesure suivants :

$$\tilde{G} \stackrel{\text{def}}{=} \bigotimes_{i\in\mathcal{N}} \tilde{G}_i, \qquad \tilde{\mu} \stackrel{\text{def}}{=} \bigotimes_{i\in\mathcal{N}} \tilde{\mu}_i.$$

6. Auteurs : G. Fayolle et J.-M. Lasgouttes

Si les nœuds du système sont suffisamment indépendants, il est loisible d'espérer que les marginales de $\tilde{\mu}$ concernant un nombre de nœuds *petit* devant N soient proches de celles de μ.

Soit $S \subset \mathcal{N}$ un sous ensemble quelconque des nœuds de $\mathcal{X}$, de cardinal $|S|$. Pour toute fonction test $f_S \in \mathcal{B}(E^S)$, l'équation (2.6.2) implique

$$\mathbb{E}^{\tilde{\mu}} G f_S(X) = \mathbb{E}^{\tilde{\mu}} (G - \tilde{G}) f_S(X).$$

Comme les générateurs G et $\tilde{G}$ ont des structures proches, les calculs se simplifient en général suffisamment pour conduire à une borne de la forme

$$\left| \mathbb{E}^{\tilde{\mu}} G f_S(X) \right| \leq \varepsilon_S \|f_S\|_\infty, \tag{2.6.3}$$

où ε_S est petit pour un *grand* système vérifiant des conditions de mélange adéquates, qui sont pour l'essentiel des conditions de flux.

Arguments de preuve pour S fini Le lecteur pourrait sans doute être tenté d'évoquer les fameux théorèmes de Kolmogorov ou de Tulcea, qui montrent comment construire la loi d'un processus à partir de distributions *compatibles* sur les ensembles cylindriques (i.e. des marginales). Ici la situation est cependant bien différente. car les processus $X^{(\mathcal{N})}(t)$ ne sont a priori aucunement comparables lorsque $\mathcal{N}$ varie. En outre, comme il ressort de l'équation (2.6.3), il s'agit d'un problème d'*approximation* de mesure.

Pour prouver que les composantes de $\tilde{\mu}$ sur S sont proches de celles de μ (au sens d'une distance classique), l'idée est de recourir à l'analyse spectrale. D'abord, on remarque que S, à une renumérotation près, agit comme une fenêtre lorsque $\mathcal{N}$ varie. Alors, le point saillant est de vérifier que l'opérateur de projection de G sur l'ensemble $S \subset \mathcal{N}$ admet la valeur propre 1 comme point isolé de son spectre (sur un espace de fonctions convenablement choisi) : or cette propriété semble vraie pour une classe assez large de *bons* systèmes, incluant la plupart des réseaux usuels (files d'attente ou assimilés).

2.6.2 Application à un système avec backbone

On considère un modèle simple (figure 2.6.1) dans lequel un backbone (lien 0) est relié à des liens « utilisateurs » (correspondant à l'ensemble noté $\mathcal{N}$). Des connexions sont créées sur chaque route, comprenant le backbone et un lien $i \in \mathcal{N}$, selon un processus de Poisson d'intensité λ_i. Chaque connexion i est maintenue jusqu'à transmission complète de ses données, dont le volume suit une loi exponentielle de moyenne σ_i.

Le nombre de connexions actives sur le lien 0 est $X_0 = \sum_{i \in \mathcal{N}} X_i$. On mélangera les notions de route ou de lien sortant, qui sont en correspondance bi-univoque. Chaque lien $i \in \mathcal{N} \cup \{0\}$ a une bande passante C_i et sa charge est définie par

$$\begin{aligned} \rho_i &\stackrel{\text{def}}{=} \frac{\lambda_i \sigma_i}{C_i}, \ i \in \mathcal{N}, \\ \rho_0 &\stackrel{\text{def}}{=} \frac{1}{C_0} \sum_{i \in \mathcal{N}} \lambda_i \sigma_i , \end{aligned}$$

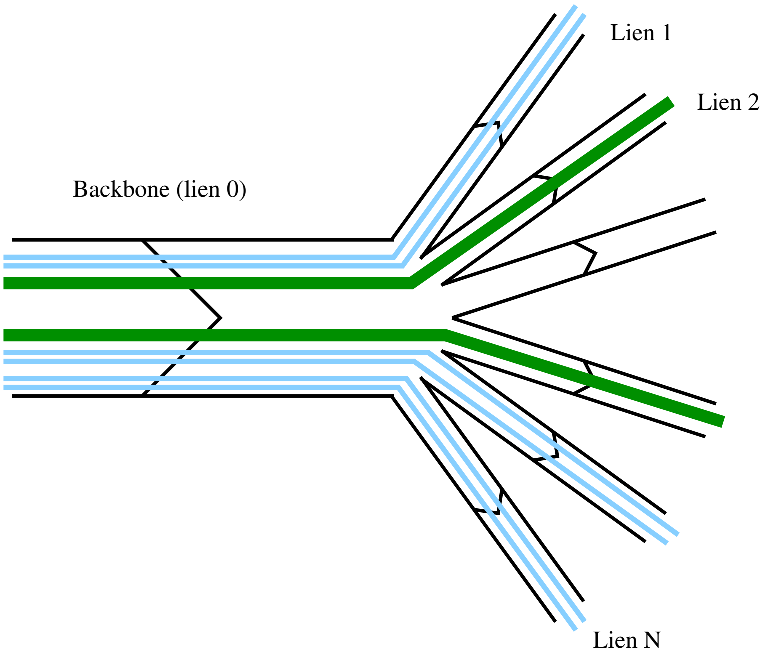

FIG. 2.6.1: *réseau asymétrique avec un seul backbone.*

avec $\rho_i < 1, \forall i \geq 0$. La bande passante allouée à la route passant par le lien $i \in \mathcal{N}$ est notée $\zeta_i(X)$.

Avec les notations de la section précédente, on peut écrire

$$\begin{aligned} Gf(X) &= \sum_{i \in \mathcal{N}} \lambda_i \big[f(X+e_i) - f(X)\big] \\ &\quad + \sum_{i \in \mathcal{N}} \frac{X_i}{\sigma_i} \zeta_i(X) \big[f(X-e_i) - f(X)\big], \end{aligned}$$

où e_i est le i-ème vecteur de base canonique de $\mathbb{R}^N$.

Dans le cas où les charges ρ_i des liens sont bien réparties (en un sens défini plus tard, mais qui inclut le cas symétrique), on peut supposer, comme dans les études précédentes, que les X_i, $i \in \mathcal{N}$ sont suffisamment indépendants quand N est grand pour que X_0 vérifie après normalisation convenable une loi des grands nombre.

Le lemme suivant sera utile pour évaluer ε_S, en notant la parenté des résultats avec les conditions de *négligeabilité asymptotique uniforme* qui apparaissent dans les théorèmes centraux limites (cf. par exemple [7]).

Lemme 2.6.1. *Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes et positives et $S_n \stackrel{\text{def}}{=} \sum_{i=1}^n X_i$. Alors, en posant $M_n \stackrel{\text{def}}{=} \mathbb{E} S_n$ et $D_n^2 \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbb{E} X_i^2$,*

$$\mathbb{P}(S_n < s) \leq \exp\Big[-\frac{(M_n - s)^2}{2D_n^2}\Big], \quad \forall s \leq M_n. \tag{2.6.4}$$

Si, de plus, pour tout i, il existe des nombres X_i^* tels que $X_i \leq X_i^*$, alors pour tout $s \geq M_n$,

$$\mathbb{P}(S_n > s) \leq \exp\Big[-\frac{(M_n - s)^2}{2D_n^{*2}}\Big], \quad \forall s \geq M_n, \tag{2.6.5}$$

où $D_n^{*2} \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbb{E}(X_i^* - X_i)^2$.

Preuve. On choisit $t > 0$ et on écrit

$$\mathbb{P}(S_n < s) = \mathbb{P}(e^{-tS_n} > e^{-ts}) \leq e^{ts}\mathbb{E}e^{-tS_n},$$

où la dernière expression provient de la célèbre inégalité de Tchebyshev. Pour chaque variable X_i, on a

$$\mathbb{E}e^{-tX_i} \leq 1 - t\mathbb{E}X_i + \frac{t^2}{2}\mathbb{E}X_i^2 \leq \exp\Big[-t\mathbb{E}X_i + \frac{t^2}{2}\mathbb{E}X_i^2\Big].$$

En prenant en compte l'indépendance des X_i, et en optimisant sur la valeur de $t > 0$, on obtient donc

$$\mathbb{P}(S_n < s) \leq \exp\Big[-t(M_n - s) + \frac{t^2}{2}D_n^2\Big] \leq \exp\Big[-\frac{(M_n - s)^2}{2D_n^2}\Big],$$

ce qui conclut la preuve de (2.6.4). Pour (2.6.5), on remarque que

$$\mathbb{P}(S_n > s) = \mathbb{P}\Big[\sum_{i=1}^n X_i^* - S_n < \sum_{i=1}^n X_i^* - s\Big],$$

et on applique (2.6.4), puisque les $X_i^* - X_i$ sont des variables positives. ■

2.6.2.1 Cas de la politique min

Le débit de chaque lien $i \in \mathcal{N}$ est calculé selon la politique min, et vaut

$$\zeta_i(X) \stackrel{\text{def}}{=} \frac{X_i}{\sigma_i} \min\Big(\frac{C_0}{X_0}, \frac{C_i}{X_i}\Big).$$

Il est alors naturel de poser, pour m_0 réel fixé,

$$\tilde{G}_i f(X_i) \stackrel{\text{def}}{=} \lambda_i\big[f(X + e_i) - f(X)\big] + \frac{X_i}{\sigma_i} \min\Big(\frac{C_0}{m_0}, \frac{C_i}{X_i}\Big)\big[f(X - e_i) - f(X)\big]. \tag{2.6.6}$$

Le choix de m_0 est fait en définissant $m_i = \mathbb{E}^{\tilde{\mu}_i}X_i$ et en posant ensuite $m_0 = \sum_{i\in\mathcal{N}} m_i$. La forme de $\tilde{G}$ montre qu'il est toujours possible de trouver de tels nombres.

Avec ces notations,

$$\begin{aligned}&(G - \tilde{G})f(X) \\ &\quad = \sum_{i\in\mathcal{N}} \frac{X_i}{\sigma_i}\Big[\min\Big(\frac{C_0}{X_0}, \frac{C_i}{X_i}\Big) - \min\Big(\frac{C_0}{m_0}, \frac{C_i}{X_i}\Big)\Big]\big[f(X - e_i) - f(X)\big].\end{aligned}$$

Soit, comme dans la section précédente, $S \subset \mathcal{N}$ et $f_S \in \mathcal{B}(E^S)$. Alors,

$$\begin{aligned}
\left|\mathbb{E}^{\tilde{\mu}}(G-\tilde{G})f_S(X)\right| &\leq 2\|f_S\|_\infty \sum_{i\in S} \mathbb{E}^{\tilde{\mu}} \frac{X_i}{\sigma_i}\left|\min\Big(\frac{C_0}{X_0},\frac{C_i}{X_i}\Big)-\min\Big(\frac{C_0}{m_0},\frac{C_i}{X_i}\Big)\right| \\
&\leq 2\|f_S\|_\infty \sum_{i\in S} \mathbb{E}^{\tilde{\mu}} \frac{X_i}{\sigma_i}\left|\frac{C_0}{X_0}-\frac{C_0}{m_0}\right| 1\!\!1_{\left\{\frac{C_i}{X_i}>\min(\frac{C_0}{X_0},\frac{C_0}{m_0})\right\}} &(2.6.7)\\
&\leq 2\|f_S\|_\infty \sum_{i\in S} \frac{C_0}{m_0\sigma_i}\mathbb{E}^{\tilde{\mu}}\Big[X_i\Big|\frac{m_0}{X_0}-1\Big|\Big] &(2.6.8)\\
&\stackrel{\text{def}}{=} \varepsilon_S\|f_S\|_\infty, &(2.6.9)
\end{aligned}$$

où (2.6.8) est relativement brutale, car elle ne prend pas en compte l'indicatrice présente dans (2.6.7). Il est donc possible, mais techniquement délicat, d'affiner la valeur de ε_S.

Posons $S_{[i]} = X_0 - X_i$ et $M_{[i]} = m_0 - m_i$. Alors, en utilisant le lemme 2.6.1, pour $\alpha \in]0,1[$ fixé,

$$\begin{aligned}
&\mathbb{E}^{\tilde{\mu}}\Big[X_i\Big|\frac{m_0}{X_0}-1\Big| \;\Big|\; X_i=x\Big] \\
&\quad= x\mathbb{E}^{\tilde{\mu}}\Big|\frac{M_{[i]}+m_i}{S_{[i]}+x}-1\Big| \\
&\quad= x\mathbb{E}^{\tilde{\mu}}\Big[\Big|\frac{M_{[i]}+m_i}{S_{[i]}+x}-1\Big|1\!\!1_{\{S_{[i]}<\alpha M_{[i]}\}}\Big] + x\mathbb{E}^{\tilde{\mu}}\Big[\Big|\frac{M_{[i]}+m_i}{S_{[i]}+x}-1\Big|1\!\!1_{\{S_{[i]}\geq\alpha M_{[i]}\}}\Big] \\
&\quad\leq (|x-m_i|+M_{[i]})\exp\Big[-\frac{(1-\alpha)^2M_{[i]}^2}{2D_{[i]}^2}\Big] + x\frac{|x-m_i|+(1-\alpha)M_{[i]}}{\alpha M_{[i]}}.
\end{aligned}$$

On choisit alors

$$(1-\alpha)^2 = \frac{2D_{[i]}^2}{M_{[i]}^2}\log M_{[i]}^2$$

et on obtient

$$\mathbb{E}^{\tilde{\mu}}\Big[X_i\Big|\frac{m_0}{X_0}-1\Big| \;\Big|\; X_i=x\Big] \leq \frac{|x-m_i|}{M_{[i]}^2} + \frac{1}{M_{[i]}} + x\frac{|x-m_i|}{M_{[i]}} + \frac{2D_{[i]}}{M_{[i]}\alpha}\sqrt{\log M_{[i]}}$$

Même si cette majoration est explicite et permet de tout exprimer a partir des deux premiers moments, elle n'est pas très simple à manipuler. Notons toutefois que, quand les ρ_i sont suffisamment loin de 1 pour que les variables X_i aient des moments uniformément bornés, il est possible (comme dans [7]), de montrer

$$\varepsilon_S = O\left(|S|\frac{C_0\sqrt{\log m_0}}{m_0^{3/2}}\right)$$

Il est intéressant de voir sous quelles conditions a lieu un phénomène d'insensibilité de la distribution des X_i par rapport aux paramètres du réseau (en l'occurrence C_0 et les autres ρ_i). De l'équation (2.6.6), il ressort que le nœud i sera distribué comme une file M/M/1 de paramètre ρ_i dès que $C_0 > C_i m_0$.

La zone de paramètres dans laquelle toutes les files sont insensibles à C_0 est donc décrite par

$$\max_{i\in\mathcal{N}} C_i < \frac{C_0}{m_0}. \tag{2.6.10}$$

Ce domaine est assez compliqué à décrire, puisque m_0 dépend directement des C_i au travers de ρ_i. Rappelons que, dans le cas symétrique ($\rho_i \equiv \rho$ pour tout $i \in \mathcal{N}$), la condition (2.6.10) se réduit à $\rho_0 + \rho < 1$.

2.6.2.2 Cas d'une politique efficace

On considère une politique d'allocation efficace, au sens de la définition 1.1.3. Un cas particulier notable d'une telle politique est la politique max-min. Il est intéressant de noter toutefois que la structure exacte de la politique n'est pas nécessaire pour obtenir des résultats d'insensibilité.

On considère l'allocation $\tilde{\zeta}_i$ qui serait celle du lien i si on avait $C_0 = \infty$. D'après (1.1.4), on a $\tilde{\zeta}_i(X) = C_i \mathbb{1}_{\{X_i>0\}}$. La bande passante totale réclamée dans le lien 0 est donc

$$S_N = \sum_{i\in\mathcal{N}} C_i \mathbb{1}_{\{X_i>0\}}.$$

Au vu de (1.1.5) appliquée pour $\ell = 0$, on aura $\zeta_i = \tilde{\zeta}_i$ pour tous les liens dès que $S_N < C_0$, ce qui devrait être souvent, vu que $\mathbb{E}S_n = \rho_0 C_0$. Dans ce cas, chaque route se comportera comme une file $M/M/1$; il est donc naturel de poser

$$\tilde{G}_i f(X_i) \overset{\text{def}}{=} \lambda_i\big[f(X+e_i) - f(X)\big] + \frac{C_i}{\sigma_i}\mathbb{1}_{\{X_i>0\}}\big[f(X-e_i) - f(X)\big].$$

On a alors

$$(\tilde{G} - \tilde{G})f(X) = \sum_{i\in\mathcal{N}} \mathbb{1}_{\{S_n>C_0\}}\Big[X_i\zeta_i(X) - \frac{C_i}{\sigma_i}\mathbb{1}_{\{X_i>0\}}\Big]\big[f(X-e_i) - f(X)\big]$$

et, d'après le lemme 2.6.1,

$$\big|\mathbb{E}^{\tilde{\mu}}(\tilde{G} - \tilde{G})f(X)\big| \;\le\; 2\|f\|_\infty \exp\left[-\frac{(1-\rho_0)^2 C_0^2}{2\sum_{i\in\mathcal{N}}(1-\rho_i)C_i^2}\right]\sum_{i\in\mathcal{N}}\frac{C_i}{\rho_i} \overset{\text{def}}{=} \varepsilon\|f\|_\infty.$$

La première remarque à propos de cette inégalité est qu'elle est tellement forte que la convergence vers un modèle produit semble être vraie pour la loi jointe de *toutes* les files. Dans ce cas, on aura un résultat d'insensibilité, c'est-à-dire que la distribution d'un lien utilisateur ne dépend pas des paramètres des autres liens (et en particulier du backbone).

On voit apparaître dans l'exponentielle des termes liés à la bande passante résiduelle $(1-\rho)C$. En particulier, dans un cas symétrique avec les notations $C_0 \equiv Nc_0$, $C_i \equiv C$, $\rho_i \equiv \rho_N$, $\rho_0 \equiv \rho_{0N}$, on obtient

$$\varepsilon = \frac{2NC}{\rho}\exp\left[-\frac{N(1-\rho_{0N})^2 c_0^2}{C^2(1-\rho_N)}\right].$$

On voit que, dans le cas où le le backbone est chargé ($\rho_{0N} \to 1$), il faut au pire avoir $1 - \rho_{0N} \approx 1/\sqrt{N}$, sauf si les liens utilisateurs tendent eux aussi a être saturés.

2.6.3 Un réseau de Jackson avec des temps de service non exponentiels

On exhibe ici un réseau pour lequel la méthode de la section 2.6.1 conduit à des calculs de nature différente. Il s'agit d'un réseau de Jackson à N nœuds, avec durées de service i.i.d. arbitraires à chaque nœud, la politique de service étant *premier arrivé-premier servi*. Ce réseau n'a bien sûr pas les bonnes propriétés de forme produit de la version avec temps de service exponentiels. Toutefois, sous certaines conditions, il semble possible d'approximer les marginales en limite thermodynamique.

L'état à l'instant $t \geq 0$ du nœud $i \in \mathcal{N}$ est donné par le nombre de clients présents $X_i(t)$ et le temps de service déjà reçu $R_i(t)$ par le client en cours dans la file. Par convention, on prend $R_i(t) = 0$ quand $X_i(t) = 0$.

La dynamique du réseau est décrite par les règles suivantes :

- le nœud i reçoit un flux externe poissonnien de clients d'intensité λ_{0i} ;
- les clients quittent le nœud i avec intensité $\gamma_i(X_i(t), R_i(t))$;
- quand un client quitte le nœud i, il quitte le réseau avec probabilité p_{i0}, ou rejoint le nœud $j \in \mathcal{N}$, avec probabilité p_{ij} (avec $p_{i0} + \sum_{j\in\mathcal{N}} p_{ij} = 1$).

Si on pose $E \stackrel{\text{def}}{=} \mathbb{N} \times \mathbb{R}^+$, le réseau est alors décrit par le processus de Markov

$$(X(t), R(t))_{t\geq 0} = (X_i(t), R_i(t), i \in \mathcal{N})_{t\geq 0}, \tag{2.6.11}$$

à valeurs dans E^N et qui sera supposé ergodique. Son générateur s'écrit

$$\begin{aligned} Gf(X,R) &= \sum_{i\in\mathcal{N}} \lambda_{0i}[f(\mathsf{T}_{0i}X, R) - f(X,R)] \\ &\quad + \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}\cup\{0\}} p_{ij}\gamma_i(X_i,R_i)[f(\mathsf{T}_{ij}X, \mathsf{Z}_iR) - f(X,R)] \\ &\quad + \sum_{i\in\mathcal{N}} \frac{\partial f}{\partial r_i}(X,R)\mathbb{1}_{\{X_i>0\}}, \end{aligned}$$

où $f \in \mathcal{B}(E^N)$ et où les opérateurs T et Z sont définis, $\forall i,j \in \mathcal{N}$, par

$$\begin{aligned} \mathsf{T}_{ij}X &\stackrel{\text{def}}{=} X - \mathbb{1}_{\{X_i>0\}}e_i + e_j, \\ \mathsf{T}_{0i}X &\stackrel{\text{def}}{=} X + e_i, \\ \mathsf{T}_{i0}X &\stackrel{\text{def}}{=} X - \mathbb{1}_{\{X_i>0\}}e_i, \\ \mathsf{Z}_iR &\stackrel{\text{def}}{=} R - R_ie_i. \end{aligned}$$

Soit $(\bar\lambda_i, i \in \mathcal{N})$ la solution (unique) du système linéaire d'équations

$$\bar\lambda_j = \lambda_{0j} + \sum_{i\in\mathcal{N}} p_{ij}\bar\lambda_i, \quad j \in \mathcal{N}.$$

Pour tout $i \in \mathcal{N}$, $\forall f_i \in \mathcal{B}(E)$, on définit le générateur

$$\begin{aligned}
\tilde{G}_i f_i(X,R) &= \bar{\lambda}_i\big[f_i(\mathsf{T}_{0i}X,R) - f_i(X,R)\big] \\
&\quad + \gamma_i(X_i,R_i)[f_i(\mathsf{T}_{i0}X,\mathsf{Z}_iR) - f_i(X,R)] \\
&\quad + \frac{\partial f_i}{\partial r}(X,R)\mathbb{1}_{\{X_i>0\}},
\end{aligned} \tag{2.6.12}$$

qui décrit le nœud i comme une simple file d'attente avec arrivées poissonniennes d'intensité $\bar{\lambda}_i$ et taux de service $\gamma_i(X_i, R_i)$, et on note comme précédemment $\tilde{\mu}_i$ la mesure invariante associée.

Le lemme suivant sera utile plus loin

Lemme 2.6.2. *Pour toute fonction $f \in \mathcal{B}(\mathbb{N})$, on a*

$$\bar{\lambda}_i \mathbb{E}^{\tilde{\mu}_i} f(X_i) = \mathbb{E}^{\tilde{\mu}_i}\gamma_i(X_i,R_i)f(X_i-1).$$

Preuve. C'est une conséquence directe de

$$\mathbb{E}^{\tilde{\mu}}\tilde{G}_i f(X_i) = 0,$$

en utilisant le fait que f ne dépend pas de R_i. ∎

On choisit $S \in \mathcal{N}$ et on pose $S^c \stackrel{\text{def}}{=} \mathcal{N} \setminus S$. On se donne aussi une mesure de probabilité α_{S^c} sur E^{S^c} telle que, pour $i \in S^c$,

$$\mathbb{E}^{\alpha_{S^c}}\gamma_i(X_i,R_i) = \bar{\lambda}_i. \tag{2.6.13}$$

Deux exemples possibles d'une telle mesure sont

$$\begin{aligned}
\tilde{\mu}_{S^c} &\stackrel{\text{def}}{=} \bigotimes_{i\in S^c} \tilde{\mu}_i, \\
\mu_{S^c} &\stackrel{\text{def}}{=} \int_{(X_S,R_S)\in E^S} \mu(dx,dr).
\end{aligned}$$

Dans la suite, on considère que les files de S on $\tilde{\mu}_S$ comme distribution (c'est-à-dire que ce sont des files M/G/1 indépendantes), alors que les autres sont régies par α_{S^c}. Alors, en utilisant, pour $f_S \in \mathcal{B}(E^S)$, la relation

$$\begin{aligned}
0 &= \mathbb{E}^{\tilde{\mu}_S}\tilde{G}_S f_S(X,R) \\
&= \sum_{j\in S}\Big[\lambda_{0j} + \sum_{i\in\mathcal{N}}\bar{\lambda}_i p_{ij}\Big]\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,R) - f_S(X,R)\big] \\
&\quad + \sum_{i\in S}\Big[p_{i0} + \sum_{j\in\mathcal{N}} p_{ij}\Big]\mathbb{E}^{\tilde{\mu}_S}\Big[\gamma_i(X_i,R_i)\big[f_S(\mathsf{T}_{i0}X,\mathsf{Z}_iR) - f_S(X,R)\big]\Big] \\
&\quad + \sum_{i\in S}\mathbb{E}^{\tilde{\mu}_S}\frac{\partial f_S}{\partial r_i}(X,R)\mathbb{1}_{\{X_i>0\}},
\end{aligned}$$

donne, après quelques calculs,

$$\begin{aligned}
&\mathbb{E}^{\tilde{\mu}_S} G f_S(X,R) \\
&\quad = \mathbb{E}^{\tilde{\mu}_S}(G - \tilde{G}_S) f_S(X,R) \\
&\quad = \sum_{i\in S^c, j\in S} p_{ij}[\gamma_i(X_i,R_i) - \bar{\lambda}_i]\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,R) - f_S(X,R)\big] \\
&\qquad + \sum_{i,j\in S} p_{ij}\mathbb{E}^{\tilde{\mu}_S}\Big[\gamma_i(X_i,R_i)\big[f_S(\mathsf{T}_{ij}X,\mathsf{Z}_iR) - f_S(\mathsf{T}_{i0}X,\mathsf{Z}_iR)\big]\Big] \\
&\qquad - \sum_{i,j\in S} p_{ij}\bar{\lambda}_i\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,R) - f_S(X,R)\big].
\end{aligned}$$

Alors, en appliquant le lemme 2.6.2,

$$\begin{aligned}
&\mathbb{E}^{\tilde{\mu}_S} G f_S(X,R) \\
&\quad = \sum_{i\in S^c, j\in S} p_{ij}\Big[\gamma_i(X_i,R_i) - \bar{\lambda}_i\Big]\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,R) - f_S(X,R)\big] \\
&\qquad + \sum_{i,j\in S} p_{ij}\bar{\lambda}_i\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,\mathsf{Z}_iR) - f_S(X,\mathsf{Z}_iR) \\
&\qquad\qquad - f_S(\mathsf{T}_{0j}X,R) + f_S(X,R)\big],
\end{aligned}$$

et finalement, grâce à (2.6.13),

$$\begin{aligned}
\mathbb{E}^{\alpha_{S^c}\otimes\tilde{\mu}_S} G f_S(X,R) &= \mathbb{E}^{\alpha_{S^c}}\mathbb{E}^{\tilde{\mu}_S} G f_S(X,R) \\
&= \sum_{i,j\in S} p_{ij}\bar{\lambda}_i\mathbb{E}^{\tilde{\mu}_S}\big[f_S(\mathsf{T}_{0j}X,\mathsf{Z}_iR) - f_S(X,\mathsf{Z}_iR) \\
&\qquad - f_S(\mathsf{T}_{0j}X,R) + f_S(X,R)\big].
\end{aligned}$$

On obtient donc une inégalité de type (2.6.3) avec

$$\varepsilon_S \stackrel{\text{def}}{=} 4\sum_{i,j\in S} p_{ij}\bar{\lambda}_i \le 4|S|^2 \sup_{i,j\in\mathcal{N}} p_{ij}\bar{\lambda}_i,$$

grandeur qui est faible en particulier dans le cas d'un réseau symétrique complètement connecté.

Table des figures

Bibliographie

[1] D. Bertsekas et R. Gallager. *Data Networks*. Prentice-Hall International, 2nd édition, 1992.

[2] H. Blasius. « Grenzschichten in Flüssigkeiten mit kleiner Reibung ». *Z. Math Phys.*, 56:1–37, 1908. English translation in NACA TM 1256.

[3] F. Delcoigne et A. de La Fortelle. « Large Deviations Problems for Star Networks: the Min Policy ». Rapport de recherche 4143, INRIA, 2001.

[4] F. Delcoigne et G. Fayolle. « Thermodynamical limit and propagation of chaos in polling systems ». *Markov Processes and Related Fields*, 5(1):89–124, 1999.

[5] G. Doetsch. *Handbuch der Laplace Transformation*, volume 1-3. Birkhaüser-Verlag. Basel, 1955.

[6] G. Fayolle, A. de La Fortelle, J.-M. Lasgouttes, L. Massoulié, et J. Roberts. « Best-effort networks: modeling and performance analysis via large networks asymptotics ». Dans *Proceedings of IEEE INFOCOM' 2001*, 2001.

[7] G. Fayolle et J.-M. Lasgouttes. « Asymptotics and scalings for large closed product-form networks via the Central Limit Theorem ». *Markov Processes and Related Fields*, 2(2):317–348, 1996.

[8] G. Fayolle et J.-M. Lasgouttes. A nonlinear integral operator encountered in the bandwidth sharing of a star-shaped network. Dans *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pages 231–242. Birkhaüser, 2000.

[9] G. Fayolle, V. A. Malyshev, et M. V. Menshikov. *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge University Press, 1995.

[10] F. P. Kelly. « Charging and rate control for elastice traffic ». *Europ. Trans. Telecom.*, 8:33–37, 1997.

[11] L. Massoulié et J. Roberts. Communication personnelle, 1998.

[12] J. Roberts et L. Massoulié. « Bandwidth sharing and admission control for elastic traffic ». ITC specialists seminar, 1998.

[13] H. Schlichting. *Boundary layer theory*. McGraw-Hill Book Co., Inc., New York, 1960. Translated by J. Kestin. 4th ed. McGraw-Hill Series in Mechanical Engineering.

[14] J. Steinheuer. « Die Lösungen der Blasiusschen Grenzschichtdifferentialgleichung ». *Abh. Braunschweig. Wiss. Gesell.*, 20:96–125, 1968.

[15] E. C. Titchmarsh. *The Theory of Functions.* Oxford University Press, 2nd édition, 1939.

[16] E. C. Titchmarsh. *Introduction to the Theory of Fourier Integrals.* Oxford University Press, 2nd édition, 1948.

[17] G. De Veciana, T.-J. Lee, et T. Konstantopoulos. « Stability and performance analysis of networks supporting services with rate control – could the Internet be unstable? ». Dans *Proceedings of IEEE Infocom'99*, pages 802–810, 1999.

[18] N. D. Vvedenskaya, R. L. Dobrushin, et F. I. Karpelevich. « A queueing system with a choice of the shorter of two queues—an asymptotic approach ». *Problems Inform. Transmission*, 32:15–27, 1996.

Notations

Notations générales

$\mathbb{R}$, ensemble des nombres réels ;

$\mathbb{Z}$, ensemble des entiers relatifs ;

$\mathbb{N}$, ensemble des entiers naturels ;

$|A|$, cardinal de l'ensemble A ;

A^c, complémentaire de l'ensemble A ;

$\mathbb{E}(X)$, espérance de la variable aléatoire X ;

$\mathbb{E}^{\mu}(X)$, espérance de la variable aléatoire X sous la loi μ ;

$\mathbb{P}(E)$, probabilité de l'événement E.

Notations propres au modèle

$\mathcal{L}$, ensemble des liens formant le réseau ;

N, nombre de liens dans le réseau ($N = |\mathcal{L}|$) ;

$\mathcal{R}$, ensemble des routes, chacune notée $r = (r(1), \ldots, r(L_r)) \in \mathcal{L}^{L_r}$, où L_r est la longueur de $r \in \mathcal{R}$;

$\ell \in r$, « le lien $\ell \in \mathcal{L}$ fait partie de la route $r \in \mathcal{R}$ »;

$r \ni \ell$, « la route $r \in \mathcal{R}$ passe par le lien $\ell \in \mathcal{L}$ »;

λ_r, taux de création de connexions sur la route $r \in \mathcal{R}$;

σ_r, volume moyen de données transmises sur une route $r \in \mathcal{R}$;

C_ℓ, capacité du lien $\ell \in \mathcal{L}$;

ρ_ℓ, charge totale du lien $\ell \in \mathcal{L}$;

$x_r(t)$, nombre de connexions actives sur la route $r \in \mathcal{R}$ à l'instant t ;

$X_\ell(t)$, nombre de connexions actives sur le lien $\ell \in \mathcal{L}$ à l'instant t ;

$\zeta_r(X)$, capacité allouée à la route $r \in \mathcal{R}$ quand le réseau est dans l'état X.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399