N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBES

THÈSE
DE DOCTORAT

Présentée par

SAKHI HANANE

Spécialité : *Mathématiques*

Option : probabilités- *Statistique*

Intitulée

Analyse d'un réseau fluide multi-classe

Soutenue le : 2016 /2017

Devant le jury composé de :

Président :	***Abderrahmane YOUSFATE***	***Univ. Djillali Liabes SBA (Pr)***
Examinateurs :		
	Toufik GUENDOUZI	***Univ. Moulay Tahar Saida (Pr)***
	Abdeldjebbar KANDOUCI	***Univ. Moulay Tahar Saida (MCA)***
	Fethi MADANI	***Univ. Moulay Tahar Saida (MCA)***
	Abbes RABHI	***Univ. Djillali Liabes SBA (MCA)***
Directeur de thèse:	***Amina Angelika BOUCHENTOUF***	***Univ. Djillali Liabes SBA (MCA)***

Dédicace

Je dédie ce modeste travail à ceux que je possède de plus cher au monde, qui depuis toujours portent un grand intérêt pour mes études, merci mes parents pour votre amour, éducation, encouragement, sacrifices et votre soutien que vous m'avez assuré tout au long de mes études et ma vie.

A mes soeurs Nabila et Achwak, mon frére Kadda.

A mon mari fethi celui qui m'a aidé pendant la réalisation de ce travail.

A ceux qui j'ai passé avec eux les meilleurs moments de ma vie.

A tout ma famille et mes collégues que je ne peux tous les citer et avec qui je partage de nobles sentiments.

Remerceiments

Je remercie d'abord **ALLAH** le tout puissant de nous avoir donné la force, la patience et la volonté pour achever ce travail.
Je tiens à remercier mon encadreur Dr. Amina Angelika BOUCHENTOUF d'avoir dirigé cette thèse, pour sa contribution scientifique, et pour l'intérêt permanent qu'elle a manifesté tout au long de ce travail, ses conseils et remarques qui m'ont facilité le travail.
Je tiens également à remercier les membres de jury : Au Pr Abderrahmane YOUSFATE qui me fait l'honneur de présider ce jury et aux professeurs Abbes RABHI, Abdeldjebbar KANDOUCI, Fethi MADANI, et Toufik GUENDOUZI et qui m'ont fait l'honneur d'examiner ce travail.
Je ne saurais manquer de rendre hommage à mes chers parents qui n'ont cessé de ménager leurs aides et leurs encouragements.
Je souhaiterais remercier tous mes amis qui m'ont aidé, encourager et soutenus pendant toutes la durée de l'élaboration de cette thèse.
Enfin, je remercie toutes personnes ayant contribué de prés ou de loin à la réalisation de ce travail.

Table des matières

Résumé

Dans cette thèse nous étudions la stabilité de différents réseaux de files d'attente fluides multiclasses. En utilisant la fonction de Lyapunov linéaire par morceaux, nous établissons une condition suffisante pour assurer la stabilité globale d'un système de files d'attente multiclasse composé de N stations et N^2 classes de clients sous des disciplines de services avec prioirité. La détermination de la région de la stabilité globale d'un tel système est trés difficile sous certaines disciplines de service, même si l'intensité du trafic à chaque unité du réseau est moins que 1. Ensuite nous étudions la stabilisation d'un modèle de réseaux de files d'attente composé de N stations, $N \geq 3$ et $2N$ classes de clients sous des disciplines de service avec priorité. En se basant sur le modèle fluide la stabilité du réseau est établie. Et nous terminons cette thèse par l'analyse d'un système de centre d'appel composé de deux orbites réservées aux clients impatients dites "redial orbits", et deux autres réservées aux clients revenant de l'éxterieur appelées "reconnect orbits", s serveurs et deux flux de clients independants suivant un processus de Poisson. Dans ce travail, le modèle fluide est utilisé pour calculer une approximation du premier ordre du nombre de clients dans les orbites du système.

Abstract

In this thesis we study the stability of various fluid multiclass queueing system.

At first, we establish a sufficient condition to ensure global stability for queueing network composed of N-units, $N \geq 3$ and N^2 classes, under priority service discipline. Determining the global stability region of such system is very difficult under bad disciplines, even if the traffic intensity at each unit of the network is less that one. Then, we establish the stability of fluid queueing network model under priority service discipline. To this end, we introduce a priority fluid multiclass queueing network model, composed of N stations, $N \geq 3$ and $2N$ classes. In this work the fluid model approach is employed in the study of the stability. Finally we analyze a call center system with two reconnect orbits, two redial (retrial) orbits, "s" servers and two independent Poisson streams of customers. In this work the fluid model is used to derive a first order approximation for the number of customers in the redial and reconnect orbits.

Chapitre 1

Introduction et Présentation

L'étude formelle des phénomènes d'attente a commencé au début du XXe siècle avec les travaux de A.K. Erlang (1917). Cet ingénieur danois a fourni la première analyse mathématique d'un modèle de files d'attente pour un commutateur téléphonique. Depuis lors, les nouvelles technologies ont constamment favorisé la nécessité d'accroître la recherche sur ces phénomènes. Les développements dans les systèmes de fabrication, d'informatiques et de communication au cours de la seconde moitié du XXe siècle ont été motivés, en particulier, l'étude des systèmes interconnectés sous la forme de réseaux de files d'attente. L'utilité des réseaux de files d'attente est devenu claire avec le travail de Scherr dans les années 1960, qui a étudié les retards et les temps d'attente dans les réseaux informatiques. Dans les années 1970 de nombreux chercheurs ont commencé à utiliser les réseaux de files d'attente pour la modélisation des systèmes de fabrication flexibles. Dans les années 1980 les modèles de réseaux de files d'attente ont commencé à être utilisé pour l'analyse des réseaux de communication, et à partir de ce moment-là, l'analyse des réseaux de files d'attente a été développée de plus en plus, de sorte que la théorie des files d'attente est devenue l'un des domaines les plus importants et actifs de recherche en recherche opérationnelle et les probabilités appliquées. Les modèles de réseaux de files d'attente ont été appliqués avec succés dans l'évaluation des performances et l'optimisation des systèmes informatiques, des systèmes de communication, des systémes de fabrication et de systémes logistiques ..... Cependant, pour de nombreux réseaux de files d'attente, les questions de stabilité et de l'évaluation de la performance restent des problémes extrêmement difficiles.
Un grand intérêt a été attribué à la compréhension de la dynamique des réseaux de files d'attente multi-classe, et en particulier leurs propriétés de stabilité. De nombreuses techniques ont été développées pour l'analyse de la stabilité ou l'ergodicité en utilisant une variété de méthodes.
L'analyse de la stabilité et de l'évaluation de la performance des réseaux de files d'attente ont reçu beaucoup d'attention. Ceci est dû à plusieurs exemples qui démontrent que les conditions de stabilité habituelles (intensité du trafic inférieure à 1 à chaque station) ne sont pas suffisantes pour la stabilité, même dans le cadre de la discipline de service FIFO voir par exemple, Rybko et Stolyar, Bramson (1993). Les méthodes utilisées pour établir la stabilité des réseaux de files d'attente ont été développées par plusieurs chercheurs, en utilisant des limites fluides, voir Dai, Chen (1995), et les fonctions de Lyapunov par exemple Kumar et Meyn (1995), (1996). Les approches de programmation mathématique pour obtenir les limites des paramétres de performance pour un réseau de files d'attente supposé stable ont été simultanément mises au point par Bertsimas et al.(1996) , Kumar et Kumar (1994).

Dans cette thèse, nous nous intéressons à l'analyse de certaines catégories de réseaux fluides multiclasses. Notons que beaucoup d'efforts ont été consacrés à l'étude de la stabilité des réseaux fluides, par exemple Chen (1995) a établie des conditions nécessaires et suffisantes pour la stabilité des réseaux fluides sous la discipline de service general work-conserving. Les conditions de stabilité des réseaux fluides sous la discipline de service FIFO et sous les disciplines de service avec priorité ont été dérivées par Chen et Zhang (1997, 2000). Dai (1999) prouve la stabilité d'un modèle fluide d'une ligne réentrante sous la discipline de service LBFS (dernier client arrivé est le premier servi). Souvent, les fonctions de Lyapunov sont utilisées pour établir ces conditions. Ye et Chen (2002) ont étudié les réseaux fluides sous des disciplines de service avec priorité en utilisant les fonctions de Lyapunov linéaires par morceaux.

Ce chapitre introductif vise à présenter

— L'étude des réseaux de files d'attente multiclasses.
— L'étude de la stabilité des réseaux fluides associés

Dans la première partie de ce chapitre (Section 1), nous présentons les bases conceptuelles pour le reste de cette thése, dont le sujet principal est l'analyse de la stabilité des réseaux fluides multiclasses. Nous récapitulons sur la base des résultats remarquables développés par plusieurs chercheurs. L'analyse de la stabilité des réseaux fluides qui est un outil trés puissant pour l'investigation de la récurrence positive au sens de Harris des réseaux de files d'attente multiclasses. A cette fin, nous fournissons une description détaillée des réseaux de files d'attente multiclasses, du processus de Markov sous-jacent, et des équations de dynamique de réseaux de files d'attente qui découlent de Bramson (2008) et Dai (1995). De plus, nous discutons une approche qui vise à réduire le problème de la stabilité d'un réseau de files d'attente multiclasse à l'étude de la stabilité d'un réseau pûrement déterministe, appelé réseau fluide associé. De plus, nous recueillons des propriétés connues du réseau fluide associé qui joueront un rôle important tout au long de cette thèse.
La deuxième partie de ce chapitre (Section 2) est consacrée à l'étude de la stabilité des réseaux fluides, nous commençons par présenter des généralités sur la stabilités des réseaux fluides. Ensuite, nous étudions les critères de stabilité des réseaux fluides sous des disciplines de service particulières ; réseaux fluides sous la discipline de service general work-conserving, réseaux fluides sous des disciplines de service avec priorité, réseaux fluides sous la discipline de service FIFO. Et dans la dernière partie de cette section nous fournissons une comparaison avec la théorie de Lyapunov pour les systèmes dynamiques modélisés par des équations différentielles ordinaires.
Notons que chapitre est inspiré de la thèse de Schönlein (2012). La fin de ce chapitre introductif sera consacrée à un bref aperçu sur la contribution et le plan de cette thèse.

1.1 Réseaux de files d'attente multiclasses

Dans cette partie de thèse, nous présentons différents modèles de réseaux de files d'attente multiclasses, et nous recueillons des résultats de la littérature qui seront utils pour la suite de cette thése. L'intention de ce résumé englobe principalement deux questions.

1. Il motive pourquoi l'analyse de la stabilité des réseaux fluides est un sujet intéressant.
2. Il présente le degré de précision des résultats.

Pour la Premiére question, il est rappelé comment les réseaux fluides émergent dans l'analyse de stabilité des réseaux de files d'attente multiclasses. En outre, cette partie présente les équations de base du réseau fluide qui sont dérivées des équations du réseaux de files d'attente multiclasses, ces dérnières décrivent l'évolution des processus de performances d'un réseau de files d'attente multiclasse.

Concernant la deuxième question, le degré de précision des résultats provient du fait que les propriétés du réseau fluide indiquées dans les propositions 1.1.2, 1.1.3 et 1.1.4 jouent un rôle fondamental dans la théorie de Lyapunov pour les réseaux fluides. La description du modéle donné dans la section 1.1.1 est tirée de Bramson (2008) et Dai (1995).

Les hypothéses 1.1.1.1 sont appropriées pour établir une analyse de stabilité des réseaux de files d'attente multiclasses en utilisant le réseau fluide associé, voir Dai (1995).

En outre, la description de l'espace d'état pour les disciplines considérées dans cette partie provient de Dai (1995). Les équations de dynamique fondamentales caractérisant l'évolution du réseau de files d'attente sous certaines disciplines peuvent être trouvées dans de nombreux travaux de recherche, voir par exemple Bramson (1996), Bramson (2008), Chen (1995), Chen et Zhang (1997, 2000), Dai (1995, 1999, 2007). Le développement de l'approximation fluide est lié à Bramson (2008). En utilisant la méthode de scaling (renormalisation) des lois fortes des grands nombres des processus stochastiques (voir la section 1.1.9) évite l'étude d'un modèle fluide avec retard.

Enfin, ces résultats fournissent l'origine de l'approche des limites fluides pour aborder le probléme de stabilité. Ce cadre a été introduit par Rybko et Stolyar (1992) pour étudier la stabilité d'un réseau de files d'attente à deux stations.
Dai (1995) a généralisé cette approche aux réseaux de files d'attente multiclasses. Dans la suite, des conditions de stabilité pour les réseaux fluides associés sous des differentes disciplines sont présentées.

1.1.1 Description détaillée d'un modèle de files d'attente multiclasse

Un réseau de file d'attente se compose des entrées, une file d'attente et des serveurs comme des centres de services. En général, il se compose d'un ou plusieurs serveurs pour servir les clients qui arrivent d'une manière quelconque comportant des exigences de service. Les clients (les flux d'entités) représentent les utilisateurs, les emplois, les opérations ou programmes. Ils arrivent au centre de service, attendent pour qu'ils soient servis s'il y a une salle d'attente, et ils quittent le système à la fin du service. Parfois, les clients sont perdus. Donc, les systèmes de files d'attente sont décrits par la distribution des temps interarrivées, la distribution des temps de service, le nombre de serveurs, la discipline de service et la capacité maximale.

Notons aussi qu'un réseau de files d'attente est un ensemble de files d'attente interconnectées classé en deux catégories :
• Réseaux de files d'attente monoclasses, dans lesquels circule une seule classe de clients.
• Réseaux de files d'attente multiclasses, dans lesquels circulent plusieurs classes de clients, ces différentes

classes pouvant se distinguer par un schéma de routage spécifique et par des comportements différents au niveau de chaque station, tant au niveau du service que de l'ordonnancement de l'attente.
Dans le cas de réseaux multiclasses, si toutes les classes de clients sont des classes ouvertes, on parlera de "réseaux pûrement ouverts" et si toutes les classes de clients sont des classes fermées, on parlera de "réseaux purement fermés". Un réseau parcouru à la fois par des classes ouvertes et des classes fermées sera qualifié comme " réseau mixte".

De manière générale, un réseaux de files d'attente multiclasse se compose de J stations et K différentes classes de clients. La dynamique du réseau peut étre décrite par :
$*$ Le processus $E_k(t)$, indiquant le nombre d'arrivées extérieures dans la période de temps $[0;t]$.
$*$ Le processus de service $S_k(t)$, reflètant le nombre de clients éventuellement finies leurs services dans la classe k au cours des premières unités de temps t.

Pour plus de commodité, nous supposons que chaque catégorie de clients est servie exclusivement à une station.
$*$ L'application c, déterminant quelle classe de clients est servie à et à quelle station.
$*$ $C(j)$, indiquant l'ensemble des classes de clients qui sont servies à la station j.
$*$ Le processus de routage $R_k^l(n)$, désignant le nombre de clients à la classe k aprés avoir eu leur service à la classe l.
Les valeurs moyennes des processus de comptage E_k, S_k et R_l^k sont notées par α_k, μ_k et P_{lk} respectivement, avec α_k, μ_k et P_{lk} sont finies.
$*$ Le processus d'allocation $T_k(t)$, désignant le temps de service accumulé consacré au service des clients de classe k au temps t.
$*$ Le processus $Q_k(0)$, indiquant le niveau de fluide initial dans la classe k, et Le processus $Q_k(\cdot)$ est le niveau de fluide dans la classe k au temps t.
Les niveaux de fluides doivent satisfaire

$$Q_k(t) = Q_k(0) + E_k(t) + \sum_{l=1}^{k} R_k^l(S_l(T_l(t))) - S_k(T_k(t)).$$

Pour obtenir une description complète de la dynamique du système, des conditions supplémentaires sur Q et T dépendent des disciplines de service avec priorité doivent être considérées. Dans la réalité, un réseau de file d'attente multiclasse reste une approximation d'un véritable système.

Le temps d'inter-arrivées des clients de la classe $k \in \{1, 2, ..., K\}$ est donné par des variables aléatoires positives $a_k(n)$, avec $n = 1, 2, 3, ...$, et le temps de service des clients de classe k est donné par des variables aléatoires positives $s_k(n)$, avec $n = 1, 2, 3, ...$ Chaque classe de clients est servie exclusivement dans une station donnée.

L'application $c : \{1, 2, ..., K\} \longrightarrow \{1, ..., J\}$ détermine quelle classe de client est servie et dans quelle station.

La matrice de composition correspondante $(C)_{J\times K}$

$$C_{jk} = \begin{cases} 1, & c(k) = j \\ 0, & \text{sinon} \end{cases} \tag{1.1}$$

définie par l'ensemble

$$\{C(j) := k \in \{1, 2, ..., K\} : c(k) = j\},$$

est l'ensemble de toutes les classes de clients servies à la station j. La trajectoire d'un client de classe k qui a reçu son service à la station $c(k)$ est donnée par une variable aléatoire de Bernoulli ϕ^k de dimension K.

Plus précisement, chaque composante de $\phi^k(n)$ est égale à 0 ou à 1.
Soit e_k le $k^{ième}$ vecteur dans $\mathbb{R}^K$. Alors le client de classe k dans la station $c(k)$ aprés avoir terminé son service, il passe à la classe l si $\phi^k(n) = e_l$ ou bien il quitte le réseau si $\phi^k(n) = 0$.

Remarque 1.1.1 *Pour certains clients de classe k, le temps d'inter-arrivée peut être* $a_k(n) = \infty$ *pour tout n. D'où, le processus d'arrivée exogène est nul. La notation correspondante est la suivante*

$$\{\varepsilon := k \in \{1, 2, ..., K\} : a_k(n) < \infty, n \geq 1\}.$$

De plus, la classe (le buffer) à chaque station est supposée de capacité infinie.

Dans ce qui suit, nous présentons les hypothèses générales sur le temps d'inter-arrivées et le temps de service. Pour cela, une distribution ν de temps d'inter-arrivées est dite non bornée si pour chaque classe $k \in \varepsilon$ et pour tout $t \geq 0$ nous avons

$$P_\nu[a_k(1) \geq t] := \int_t^\infty a_k(1)\nu(ds) > 0.$$

Cette expression exprime que les temps d'intérarrivées arbitrairement grands apparaissent avec une probabilité positive. En plus, la distribution des temps d'interarrivées des clients de classe $k \in \varepsilon$ est dite stable s'il existe certains $l_k \in \{1, 2, 3, ..., \}$ et une fonction positive $q_k : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ avec $\int_0^\infty q_k(s)ds > 0$ tel que pour tout $0 \leq a < b$, on a

$$P_\nu[\sum_{i=1}^{l_k} a_k(i) \in [a, b]] \geq \int_a^b q_k(s)ds.$$

Avant de résumer les hypothèses permanentes, nous définissons le rayon spectral d'une matrice $M \in \mathbb{R}^{K\times K}$ par

$$\varrho(M) = \sup\{|\lambda| \ / \ \exists x \in \mathbb{R}^K : Ax = \lambda x\}.$$

1.1.1.1 Hypothèses

(1) Les suites $a_1, a_2, ..., a_K, s_1, s_2, ..., s_K$ et $\phi^1, \phi^2, ..., \phi^K$ sont indépendantes identiquement distribuées et mutuellement indépendantes.

(2) Les premiers moments satisfairent

$$\alpha_k := \mathbb{E}[a_k(1)]^{-1} < \infty, \quad \text{pour} \quad k \in \varepsilon,$$

$$\mu_k := \mathbb{E}[s_k(1)]^{-1} < \infty, \quad \text{pour} \quad k \in \{1, 2, ..., K\},$$

$$P_k := \mathbb{E}[\Phi^k(1)]^{-1} \geq 0, \quad \text{pour} \quad k \in \{1, 2, ..., K\},$$

et le rayon spectral de la matrice $P = (P_1.....P_K)$ est strictement inférieur à 1.
(3) Les distributions des temps d'inter-arrivées sont non-bornées et étendues.

Dans tout ce qui suit, le triplet (a, s, ϕ) est considéré comme la suite des incréments primitifs du réseau de files d'attente multiclasse.

La $l^{i^{ème}}$ composante du paramètre P_k de routage de Bernoulli ϕ^k reflète la probabilité qu'un client de classe k devient un client de la classe l. Par conséquent,

$$1 - \sum_{l=1}^{K} P_{kl},$$

représente la probabilité qu'un client de classe k quitte le réseau à la fin de son service.

Puisque la matrice de transition $P = (P_{kl})$ admet un rayon spectral strictement inférieur à 1, la série de Neumann converge, c'est à dire

$$(I + P + P^2 + ...)^T = (I - P^T)^{-1},$$

existe. En conséquence, chaque client visite prèsque sûrement seulement un nombre fini de stations avant de quitter le réseau. Dans les réseaux de files d'attente multiclasses ouverts le taux d'arrivée λ_k de clients de classe k est donné par

$$\lambda_k = \alpha_k + \sum_{l=1}^{K} P_{lk}\lambda_l.$$

Alors, la charge initiale de la station j est donnée par la somme des quotients des taux d'arrivés effectifs λ_k et le taux de service μ_k dans toutes les classes de clients présentes à la station j

$$\rho_j = \sum_{k \in C(j)} \frac{\lambda_k}{\mu_k}.$$

Et comme le rayon spectral de la matrice P est strictement inférieur à 1, la forme vectorielle de taux d'arrivée est donnée comme suit

$$\lambda = (I - P^T)^{-1}\alpha.$$

De plus, en utilisant $M = diag(\mu_1, ..., \mu_K)$, la charge de travail nominale des stations par unité de temps est représentée par le vecteur de dimension J

$$\rho = CM^{-1}\lambda. \tag{1.2}$$

Notons que pendant une longue période, il a été considéré que "la charge initiale de chaque station strictement inférieure à 1" est une condition suffisante de stabilité. Cependant, Kumar et Seidman (1990) ont présenté un réseau avec deux stations et quatre classes de clients qui est instable, même si la charge de travail initiale à chaque station est inférieure à 1. Cet exemple a inspiré un certain nombre d'exemples avec de nouvelles disciplines de service. Dans la littérature, ils sont connus comme le réseau de Lu-Kumar, le réseau de Rybko-Stolyar ou le réseau de Bramson, voir par exemple Bramson (1994 et 2008), Dai (2004) et Rybko et Stolyar (1996), respectivement.

1.1.1.2 Espace d'état

Dans cette thèse, nous nous limiterons aux réseaux de files d'attente HL(head-of-the-line)(au sein de chaque classe, les clients sont servis dans l'ordre (first in first out) (FIFO)). L'évolution du réseau de files d'attente multiclasse HL sera décrite par un processus stochastique $\{X = X(t), t \geq 0\}$ et son espace d'état correspondant désigné par $(\chi, B(\chi))$
où

$$\chi = \{x : x = (y, z) = (((k, w), u, v), z) \in (\mathbb{Z} \times \mathbb{R})^\infty \times \mathbb{R}^{|\varepsilon|} \times \mathbb{R}^K \times \mathbb{R}^K\}, \tag{1.3}$$

et χ est muni d'une métrique d. Alors $(\mathbb{Z} \times \mathbb{R})^\infty$ est l'ensemble des variables finies à valeurs dans $(\mathbb{Z} \times \mathbb{R})$, et $B(\chi)$ est la norme de Borel σ-algébre de χ induite par la métrique d. Avant d'introduire la métrique, nous considérons l'espace d'état χ. La signification de chacune de ses composantes est définie par :

• $(k, w) \in (\mathbb{Z} \times \mathbb{R})^\infty$: L'ordre global des clients du réseau est donnée par les couples de variables $(k, w) = ((k_1, w_1), (k_2, w_2), ..., (k_l, w_l))$. La première entrée $k_i \in \{1, 2, ..., K\}$ désigne la classe actuelle du $i^{éme}$ client et la seconde entrée $w_i \geq 0$ refléte le temps écoulé depuis que le client i est rentré dans la classe k_i. Notons que dans chaque classe le plus ancien client (le premier dans le couple(k, w)) c'est le client qui a la plus grande deuxième coordonnée. Dans le cas de deux ou plusieurs clients avec la deuxième coordonnée identique, l'ordre des clients est croissant par rapport à leur classe. Le nombre de clients de chaque classe est désigné par

$$q = (q_1, ..., q_K),$$

et

$$\| q \|_1 = \sum_{k=1}^{K} q_k,$$

est le nombre total de clients dans l'espace d'état.

• $u \in \mathbb{R}^{|\varepsilon|}$: cette composante représente le temps d'intérarrivées résiduel, c'est à dire, la coordonnée $u_k > 0$ est le temps restant avant la prochaine arrivée d'un client de classe $k \in \varepsilon$ de l'extérieur du réseau.

• $\nu \in \mathbb{R}^K$: cette composante représente le temps de service résiduel, c'est à dire, la coordonnée ν_k désigne le temps de service restant pour le client le plus ancien de la classe k, où $\nu_k \geq 0$ et $\nu_k = 0$ seulement si $q_k = 0$.

• $z \in [0,1]^k$: la composante z_k désigne la proportion de service de la station $c(k)$ que reçoit la plus ancienne classe k de client, pendant que d'autres clients de la classe k ne reçoivent pas de service. Cela représente la propriété HL. Pour chaque station j, nous avons : Si $\sum_{k \in c(j)} q_k > 0$, alors $\sum_{k \in c(j)} z_k = 1$, où $z_k = 0 \quad \text{si} \quad q_k = 0$. Si la station j est vide, c'est à dire $\sum_{k \in c(j)} q_k = 0$ alors $\sum_{k \in c(j)} z_k = 0$.

Maintenant, nous introduisons une métrique d sur χ qui résume la différence de chaque composante. C'est à dire, pour $x, \quad \grave{x} \in \chi$. Nous définissons

$$d(x, \grave{x}) := \sum_{i=1}^{\infty} \min\{|k_i - \grave{k}_i| + |w_i - \grave{w}_i| + |z_i - \grave{z}_i|, 1\} + \sum_{k \in \varepsilon} |u_k - \grave{u}_k| + \sum_{k=1}^{K} |\nu_k - \grave{\nu}_k|.$$

L'espace métrique (d, χ) possède les propriétés suivantes.

Proposition 1.1.1 (Bramson (2008)) *L'espace métrique (χ, d) est séparable et localement compact. En particulier, χ est un espace Lusinien.*

Remarque 1.1.2 *L'espace métrique (χ, d) n'est pas complet. Une métrique complete peut être obtenue en ajoutant un terme approprié dans chacun des deuxième et troisième somme sur d, voir Bramson (2008) Section 4.1.*

Alors, l'espace d'état est mésurable. Avec un léger abus de notation, nous appelons

$$|x| := ||q|| + ||u|| + ||\nu||,$$

la norme sur χ. En outre, χ est muni de la topologie naturelle induite et $\{x \in \chi : |x| \leq \kappa\}$ est un compact de χ pour chaque $\kappa > 0$

1.1.2 Différentes disciplines de service

Dans cette partie, nous nous concentrons sur certaines disciplines populaires HL et nous considérons χ comme l'espace d'état résultant.

Nous verrons que l'espace d'état des réseaux de files d'attente spécifiques HL peut être simplifié en supprimant des informations redondantes.

De plus, une restriction de temps d'interarrivées et le temps de service permet de supprimer les coordonnées u et v des états, voir Bramson (2008).

1.1.2.1 Disciplines de service FIFO

Dans les réseaux de files d'attente sous la discipline de service FIFO (First in First out, premier entré-premier sorti (PEPS)), les clients sont servis selon le temps passé dans la file d'attente, où les clients les plus anciens sont les premiers dans le service. Soit $N_j(t) = \sum_{k \in C(j)} Q_k(t)$ la longueur de la file d'attente à la station j au temps t. Les clients à la station j sont ordonnés selon leur temps d'arrivée (premier entré-premier sorti), c'est à dire, pour $N_j(t) > 0$ nous considérons la suite

$$k_j(t) := (k_{j,1}, k_{j,2}, ..., k_{j,N_j(t)}),$$

où $k_{j,i}$ désigne le numéro du $i^{ème}$ client à la station j. Si $N_j(t) = 0$, la suite $k_j(t) = 0$.

Par conséquent, l'âge des clients peut être retiré de la description de l'espace d'état. De plus, comme le réseau de files d'attente est HL, la coordonné z dans (1.3) peut être également éliminée et l'évolution du processus de Markov peut être décrite par

$$X(t) = ((k_1(t), ..., k_J(t)), U(t), V_{k_{j,1}(t)}(t), j = 1, ..., J)).$$

Ainsi, l'espace d'état satisfait

$$\chi \subset (\mathbb{Z}_K^{\infty})^J \times \mathbb{R}_+^{|\varepsilon|+J},$$

où $\mathbb{Z}_K^{\infty}$ désigne l'ensemble des suites qui prend ces valeurs dans $\mathbb{Z}_K := \{1, 2, ..., K\}$.

1.1.2.2 Disciplines de service avec priorité

Sous des disciplines de service avec priorité chaque station ordonne les classes de clients et les sert par priorité, c'est à dire, pour chaque station j, on a l'application

$$\pi_j : C(j) \longrightarrow \{1, ..., |C(j)|\},$$

et la classe $\grave{k}$ avec $\pi_j(\grave{k}) = 1$ est dite une classe de priorité superieure. Si la station reçoit un nouveau client, le serveur sert le client de la classe de priorité superieure de la files d'attente. Ainsi, au sein de chaque classe, les clients sont servis selon l'ordre (FIFO). Si la file d'attente est vide, la station tourne au ralenti. Les disciplines de service prioritaires peuvent être distinguer en deux sous-disciplines.

1.1.2.3 Disciplines de service avec priorité et préemption

Si un client d'une classe prioritaire rentre dans la file d'attente, alors le service du premier client est arrêté immédiatement, et le client de la classe prioritaire est servi. Le serveur revient à son travail sauf s'il n'y a aucun autre client d'une classe prioritaire dans la file d'attente. En suivant les mêmes arguments que dans le cas (FIFO) le processus de Markov peut être décrit par

$$X(t) = ((Q_1(t), ..., Q_K(t)), U(t), V(t)),$$

et l'espace d'état peut être considéré comme

$$\chi \subset (\mathbb{Z}_+^K) \times \mathbb{R}_+^{|\varepsilon|+K}.$$

1.1.2.4 Disciplines de service avec priorité et non-préemption

Dans ce cas, même s'il arrive un client d'une classe de priorité plus élevée, la station doit terminer le service du client actuel. Ainsi, à chaque station j il y'a au plus un client servi à l'instant t. Alors, on pose $k_j(t)$ la classe du client servie à l'instant t, dans ce cas le processus de Markov peut être décrit par

$$X(t) = ((k_1(t), ..., k_J(t)), U(t), V_{k_{j,1}(t)}(t), j = 1, ..., J)).$$

S'il n'y a pas de clients présents à la station j au temps t, alors $k_j(t)$ tend vers zéro et $V_{k_j(t)}(t) = 0$. Par conséquent, le processus de Markov est complétement décrit par des éléments de l'espace d'état

$$\chi \subset (\mathbb{Z}_+^K) \times \mathbb{R}_+^{|\varepsilon|+K} \times \mathbb{Z}_K^J.$$

1.1.3 Processus de Markov sous-jacent

Dans cette partie, nous présentons le processus stochastique qui décrit l'évolution d'un réseau de files d'attente multiclasse. Premièrement, nous décrivons l'évolution du processus $X = \{X(t) = (Y(t), Z(t)) \in \chi; t \geq 0\}$ entre les arrivées et les départs des clients, où le processus se développe selon les taux de service constants par morceaux $Z(t)$. On notera que, sur la base de la discipline de service, la relation entre $Z(t)$ et $Y(t)$ peut être décrite par une fonction mesurable. Deuxièmement, pour décrire la progression du processus de Markov, quand il y'a une arrivée ou bien un départ des clients dans le réseau, nous considérons les incréments primitifs $\{a_k(n), s_k(n), \phi^k(n) : k \in \{1, 2, ..., K\}, n \geq 1\}$ pour construire l'évolution du X. Pour le temps t entre les arrivées et les départs, le taux de diminution des temps de services résiduels $V(t) = \nu$ est donné par $Z(t)$. Il y'a deux possibilités qui peuvent se produire :

(i) L'achévement du service : $V_k(t^-) = 0$ pour $k \in \{1, 2, ..., K\}$. La transition du client le plus ancien de classe k est donnée par ϕ^k. Par conséquent, nous avons $V_k(t) = s_k(i)$ ou $V_k(t) = 0$ si $q_k(t^-) > 0$ ou $q_k(t^-) = 0$ respectivement. Le client soit quitte le réseau ou il passe à une autre classe, $W_i(t) = w_i$

augmente avec un taux 1. En outre, les composantes des taux d'arrivées résiduels $U(t) = u$ diminuent de taux égal à 1 vers 0.

(ii) Nouvelle arrivée : $U_k(t^-) = 0$ pour $k \in \varepsilon$. Dans ce cas, un couple $(k, 0)$ est ajouté à l'état $Y(t)$ et le temps d'interarrivées résiduels est donné par $U_k(t) = a_k(i)$, où "i" représente le nombre de prochains nouveaux clients de classe k au temps t. Si le client de la classe k arrive à la file d'attente vide, c'est à dire $q_k(t^-) = 0$ alors $V_k(t^-) = 0$. Ainsi, nous avons $V_k(t) = s_k(\grave{i})$, où "$\grave{i}$" est l'indice du premier temps de service inutilisé à l'instant t.

Théorème 1.1.1 (Dai (1995)) *Le processus stochastique X est déterministe par morceau dont l'espace d'état est χ. En particulier X est un processus de Borel à droite et satisfait la propriété de Markov forte.*

1.1.4 Equations des réseaux de files d'attente

Dans cette partie, nous présentons en détail les équations qui décrivent l'évolution des processus de performances des réseaux de files d'attente multiclasses. En se basant sur les incréments primitifs (a, s, ϕ), nous définissons les cumulatifs primitifs (E, S, R) du réseau de files d'attente. Par convention, nous supposons que $E(0) = S(0) = R(0) = 0$.

$*$ $E(t)$, processus d'arrivée extèrne.

$*$ Le temps d'arrivée du $n^{ième}$ client de la classe k est donné par $a_k(1), ..., a_k(n)$.

$*$ Le processus $E_k(t)$ compte les clients de classe k arrivant de l'extérieur au temps t, il est défini par

$$E_k(t) =: \max\{n \in Z_+ : \ \ a_k(1) + ... + a_k(n) \leq t\}.$$

$*$ $S(t)$, processus de service cumulatif.

$*$ La composante $S_k(t)$ compte les achévements de service des clients de classe k dans la période de temps $[0; t]$, dans ce cas la station attribue toute sa capacité à cette classe de clients, c'est à dire

$$S_k(t) =: \max\{n \in Z_+ : s_k(1) + ... + s_k(n) \leq t\}.$$

$*$ $R(n)$, processus de routage.

Pour chaque $n \in \mathbb{Z}_+$ et pour chaque classe de clients $k \in \{1, 2, ..., K\}$ le processus de routage est défini par

$$R^k(n) := \sum_{i=1}^{n} \phi^k(i)$$

Bien entendu, les processus E, S et R sont càdlàg (continues à droite ayant une limite à gauche). Avant de définir les équations qui décrivant l'évolution du réseau de files d'attente, nous devons introduire le processus d'allocation, noté par $T := \{T(t), \ \ t \geq 0\}$. Ce processus présente le temps consacré dans une station pour servir les classes de clients présentes à cette dernière. Afin d'être précis, $T_k(t)$ désigne le temps de service accumulé $c(k)$ consacré à la station pour servir les clients de classe k dans l'intervalle $[0, t]$.

Le processus d'allocation est déterminé par une discipline de service. De la définition il s'ensuit que $T(\cdot)$ est croissant.

Etant donné le processus d'allocation T, le nombre de clients qui quittent la classe k à l'instant t est donné par $S_k(T_k(t))$.
De plus, le nombre de clients de classe l qui ont passé de la classe l vers la classe k dans la période $[0,t]$ est donné par $R_k^l(S_l(T_l(t)))$.
Ainsi, le niveau de fluide de la classe k au temps t peut être décrit par l'équation d'équilibre suivante

$$Q_k(t) = Q_k(0) + E_k(t) + \sum_{l=1}^{K} R_k^l(S_l(T_l(t))) - S_k(T_k(t)),$$

où $Q_k(0)$ désigne le nombre de clients qui se trouvent dans la file d'attente au temps $t = 0$. Les arrivées totales de clients de classe k à l'instant t sont définies par

$$A_k(t) = E_k(t) + \sum_{l=1}^{K} R_k^l(S_l(T_l(t))),$$

et les départs dans $[0,t]$ sont

$$D_k(t) := S_k(T_k(t)).$$

L'équation d'équilibre ci-dessus peut être écrite sous la forme simple suivante

$$Q_k(t) = Q_k(0) + A_k(t) - D_k(t).$$

En outre, il existe des processus additionnels qui sont utilisés pour décrire l'évolution du réseau de files d'attente multiclasse.
Le processus $W = \{W(t), \quad t \geq 0\}$ est appelé processus de la charge de travail immédiate. La composante $W_j(t)$ reflète le temps résiduel nécessaire pour servir tous les clients qui sont actuellement en attente d'être servis à la station j. En outre, pour $n = (n_1, ..., n_k)^T$, soit $\Gamma(n) = (\Gamma_1(n_1), ..., \Gamma_K(n_K))^T$ le service cumulatif défini par

$$\Gamma_k(n_k) = \sum_{i=1}^{n_k} s_k(i).$$

Ainsi, la charge de travail immédiate peut être caractérisée par

$$W(t) = C\Gamma(Q(0) + A(t)) - CT(t),$$

où C désigne la matrice composition.
En outre, le processus $I := \{I(t), \quad t \geq 0\}$ de dimension J est appelé processus de temps de repos,

ce processus désigne le temps total pendant lequel les stations ne fonctionnaient pas dans la période de temps $[0, t]$. Le processus de temps de repos peut être décrit par la condition suivante

$$I(t) = et - CT(t),$$

avec $e = (1, ..., 1)^T$. Comme $T(\cdot)$ est croissant, il s'ensuit que $I(\cdot)$ est également croissant. En se basant sur ces processus on trouve une autre propriété essentielle sur les réseaux de files d'attente multiclasses. Cette propriété est appelée la propriété "non-idling". Cela signifie qu'une station prend un temps de repose à l'instant t que si la file d'attente est vide, ou encore une station sert une classe d'un client prioritaire dès qu'elle est prête à être servie et qui ne peut pas être retardée si le serveur n'a rien d'autre à faire, on dit dans ce cas que l'ordonnoncement est non oisif c'est à dire qu'il fonctionne sans insertion de temps creux (non-idling ou work conserving).
Cela signifie que, si pour la station j, il s'avére que $I_j(t_1) < I_j(t_2)$ pour $t_1 < t_2$, alors il existe $t \in [t_1, t_2]$ tel que $W_j(t) = 0$. Comme $I(\cdot)$ est continu, la propriété non-idling peut s'écrire aussi comme

$$\int_0^\infty W_j(t) dI_j(t) = 0.$$

Finalement, avant de donner les équations des reseaux de files d'attente multiclasse, nous notons que la propriété HL d'un réseau de files d'attente peut être exprimée comme suit

$$\Gamma(D(t)) \leq T(t) < \Gamma(D(t) + e),$$

où les inégalités doivent être comprises par composante. Sous forme vectorielle, les équations d'un réseau de files d'attente multiclasse peuvent être résumées par

$$Q_k(t) = Q_k(0) + A_k(t) - D_k(t), \tag{1.4}$$

$$T(\cdot) \quad \text{est croissant, avec} \quad T(0) = 0, \tag{1.5}$$

$$W(t) = C\Gamma(Q(0) + A(t)) - CT(t), \tag{1.6}$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ est croissant}, \tag{1.7}$$

$$I_j(t) \quad \text{augmente seulement si} \quad W_j(t) = 0, \quad j \in \{1, ..., J\}, \tag{1.8}$$

$$\text{Les conditions additionnelles sur} \quad (Q(\cdot), T(\cdot)) \quad \text{sont spécifiques à la discipline.} \tag{1.9}$$

Dans ce qui suit nous nous référons à (1.4)-(1.9) comme des équations des réseaux de files d'attente.
Les processus définissant les équations de base de réseaux de files d'attente déterminent l'évolution de ce dernier. Pour cette raison, nous appelons

$$X(t) = (A(t), D(t), T(t), W(t), I(t), Q(t)),$$

le processus de réseau de files d'attente.

1.1.5 Réseaux de files d'attente sous la discipline de service FIFO

Dans les réseaux de files d'attente sous FIFO, les clients sont servis dans l'ordre de leurs arrivées. Ainsi, le processus d'allocation est déterminé par

$$D_k(t + W_j(t)) = Q_k(0) + A_k(t), \quad c(k) = j,$$

pour tout $t \geq 0$. La charge initiale est donnée par $Q(0)$ et

$$\{D_k(s) \text{ pour } s \leq W_j(0), \quad j = c(k), \quad k \in \{1, ..., K\}\}$$

1.1.6 Réseaux de files d'attente sous les disciplines de service avec priorité

Soit π la discipline de service avec priorité. Pour chaque classe k de client,

$$\Pi_k := \{\grave{k} \in \{1, ..., K\} : \quad c(\grave{k}) = c(k), \quad \pi(\grave{k}) < \pi(k)\},$$

désigne l'ensemble des classes de clients servis à une certaine station $j = c(k)$ et qui ont au moins la même pariorité que k. Par conséquent, nous définissons

$$T_k^+(t) := \sum_{l \in \Pi_k} T_l(t).$$

1.1.7 Réseaux de files d'attente sous les disciplines de service avec priorité et préemption

Pour les réseaux de files d'attente avec priorité et préemption la condition additionnelle

$$t - T_k^+(t)$$

augmente seulement si

$$Q_k^+ := \sum_{l \in Hk} Q_l(t) = 0, \quad k \in \{1, ..., K\},$$

pour tout $t \geq 0$. Cela peut ĥtre réécrit comme

$$\int_0^\infty Q_k^+(t) d(t - T_k^+(t)) = 0,$$

dont le niveau de fluide initial est donné par $Q(0)$.

1.1.8 Réseaux de files d'attente sous les disciplines de service avec priorité et non-préemption

Dans cette partie, nous présentons le processus d'allocation dans le cas de la discipline de service avec priorité et non-préemption. Pour une fonction de saut $f : \mathbb{R} \longrightarrow \mathbb{Z}_+$, le temps du dernier saut peut être

caractérisé par

$$l(t;f) = \sup\{s \leq t : \ |f((s^-) - f(s))| \geq 1\}.$$

Soit n_j le nombre de classes de clients servies à la station $j \in \{1, ..., J\}$, et soit $C(j) = \{(k_{j1}, ..., k_{jn_j})\}$, l'ensemble des classes de clients servies à la station j.
Supposons que k_{jl} est prioritaire par rapport à $k_{j,l+1}$. L'ensemble d'indicateurs pour la station j et la classe de client k_{jl} est donnée par

$$\tilde{I}_{k_{jl}}(t) = \{\exists u/ \ \ l(t, Q_{k_{jl}}) \leq u \leq t : Q_{k_{jl}}(u) = 0, \acute{l} < l, \ \ Q_{k_{jl}}(u) > 0\}.$$

où $l = (1, ..., n_j)$. De plus, en utilisant l'exposant c pour désigner le complémentaire, nous définissons les ensembles

$$I_{k_{jn_j}}(t) := \tilde{I}_{k_{jn_j}}(t).$$

$$I_{k_{jl}}(t) = \tilde{I}_{k_{jl}}(t) \cap (I_{k_{j,l+1}}(t))^c, \quad l = n_j - 1, ..., 1.$$

D'où, le processus d'allocation peut alors être défini par

$$T_{k_{jl}}(t) = \int_0^t 1_{I_{k_{jl}}(s)} ds,$$

pour $j \in \{1, ..., J\}$ et $l = (1, ..., n_j)$. La donnée initiale est définie par $Q(0)$.

1.1.9 Approximation fluide et stabilité

Dans cette partie, nous présentons une approche qui a été étudiée par Rybko et Stolyar(1992), et développée par Stolyar (1995) et Dai (1995). L'idée est basée sur l'étude des versions renormalisées (scaled) du processus de Markov. A cette fin, soit $(r_n, x_n)_{n\in\mathbb{N}}$ une suite de couples, où $r_n \in \mathbb{R}_+$ et $x_n \in \chi$ est une suite des états initiaux. On suppose que la suite de couples satisfait les conditions suivantes

$$\lim_{n\longrightarrow\infty} r_n = \infty, \ \ \limsup_{n\longrightarrow\infty} \frac{||q_n||}{r_n} < \infty, \quad \lim_{n\longrightarrow\infty} \frac{||u_n||}{r_n} = \lim_{n\longrightarrow\infty} \frac{||\nu_n||}{r_n} = 0, \tag{1.10}$$

où q_n, u_n et ν_n désignent la longueur de la file d'attente, le temps d'inter-arrivées résiduel, et le temps de service résiduel, respectivement. Dans la suite, nous considérons la famille $\acute{X} := \{X_n(t), \ \ t \geq 0, \ \ n \in \mathbb{N}\}$ des processus de Markov définie par

$$X_n(t) := \frac{1}{r_n} X^{x_n}(r_n t),$$

où l'exposant x_n exprime la dépendance à l'état initial $x_n \in \chi$. Afin d'étudier la famille $\acute{X}$ nous commençons à mettre l'accent sur les cumulatifs primitifs (E, S, R). Dans le lemme suivant nous rappelons les résultats de convergence des versions renormalisées des cumulatifs primitifs.

Lemme 1.1.1 (Dai (1995)) *On suppose que la suite de couples $(r_n, x_n)_{n\in\mathbb{N}}$ satisfait (1.10). Alors, presque sûrement, quand $n \longrightarrow \infty$,*

$$\frac{1}{r_n}E_k^{x_n}(r_n t) \longrightarrow \alpha_k t \quad c.u.c^{\,1}, \quad \frac{1}{r_n}S_k^{x_n}(r_n t) \longrightarrow \mu_k t \quad c.u.c, \quad \frac{1}{r_n}R^k([r_n t]) \longrightarrow P_k t \ c.u.c, \tag{1.11}$$

où $[a]$ désigne la partie entière de $a \in \mathbb{R}$

Théorème 1.1.2 *(Dai (1995)) Pour chaque réseau de files d'attente HL, $(r_n, x_n)_{n\in\mathbb{N}}$ satisfaisant (1.10) et $\omega \in G$ il existe une sous-suite de couples $(r_{n_i}, x_{n_i})_{i\in\mathbb{N}}$ de telle sorte que*

$$\lim_{i\longrightarrow\infty}\frac{1}{r_{n_i}}X^{x_{n_i}}(r_{n_i}t, \omega) = \overline{X}(t, \omega) \qquad c.u.c. \tag{1.12}$$

Toute limite :

$$\overline{X}(\cdot) = (\overline{A}(\cdot), \overline{D}(\cdot), \overline{T}(\cdot), \overline{W}(\cdot), \overline{I}(\cdot), \overline{Q}(\cdot))$$

obtenue à partir de l'équation (1.12) est appelée une limite fluide sous une certaine discipline de service. L'ensemble de toutes les limites fluides associées à la trajectoire ω est désignée par $\mathfrak{FL}(\omega)$. Donc, chaque fois qu'une limite fluide est considérée, nous avons $\overline{X}(\cdot) \in \mathfrak{FL}(\omega)$ pour certaine $\omega = \{a(n), s(n), \phi(n), \quad n \in \mathbb{N}\}$. Un modèle de limite fluide selon Schönlein (2012) est définit comme suit :

Définition 1.1.1 *L'ensemble de toutes les limites fluides pour toutes les trajectoires d'échantillon ω est appelé un modèle de limite fluide (fluid limit model ; $\mathfrak{FLM}$), c'est à dire, $\mathfrak{FLM} = \{X(\omega), \quad \omega \in G\}$.*

Où G est l'ensemble de tous les chemins d'échantillon satisfaisant (1.11).
Maintenant, nous montrons que toute limite fluide satisfait les équations de dynamique analogues aux équations du réseau de files d'attente (1.4)-(1.9), et qui sont obtenues en remplaçant les cumulatifs primitifs par leurs limites de scaling (renormalisation) ; en utilisant $M = diag(\mu)$

$$\overline{A}(t) = \alpha t + P^T M\overline{T}(t), \tag{1.13}$$

$$\overline{Q}(t) = \overline{Q}(0) + \overline{A}(t) - M\overline{T}(t) \geq 0, \tag{1.14}$$

$$\overline{T}(0) = 0 \quad \text{et} \quad \overline{T}(\cdot) \quad \text{est} \quad \text{croissante}, \tag{1.15}$$

1. La suite $(x_n)_{n\in\mathbb{N}}$ dans $D(\mathbb{R}_+, \mathbb{R}^N)$ (espace de Skorokhod) converge uniformement sur des ensembles compacts (c. u. c) vers $x \in D(\mathbb{R}_+, \mathbb{R}^N)$ si pour chaque $T > 0$, on a

$$\lim_{n\longrightarrow\infty}\sup_{t\in[0,T]} \| x_n(t) - x_n(0) \| = 0.$$

$$\overline{W}(t) = CM^{-1}(\overline{Q}(0) + \overline{A}(t)) - C\overline{T}(t), \tag{1.16}$$

$$\overline{I}(t) = et - C\overline{T}(t) \quad \text{et } \overline{I}(\cdot) \quad \text{est} \quad \text{croissante}, \tag{1.17}$$

$$\overline{I}_j(t) \quad \text{augmente seulement} \quad \text{si} \quad W_j(t) = 0, \quad j \in \{1, ..., J\}, \tag{1.18}$$

$$\text{Des conditions additionnelles sur } (\overline{Q}(\cdot), \overline{T}(\cdot)) \text{ sont spécifiques à la discipline.} \tag{1.19}$$

Pour plus de commodité, nous nous basons dans ce qui suit sur les équations fluides précedentes. Une solution fluide selon Schönlein (2012) est définit ainsi :

Définition 1.1.2 *Le couple $(\overline{Q}(\cdot), \overline{T}(\cdot))$ est appelé une solution fluide si elle satisfait les équations fluides. De plus, l'ensemble de toutes les solutions fluides des équations (1.13)-(1.19) est appelé le réseau fluide associé, désigné par $\mathfrak{FN}$.*

Le réseau fluide associé ($\mathfrak{FN}$; fluid network) est un réseau pûrement déterministe basé sur les valeurs moyennes des incréments primitifs du réseau de files d'attente stochastique.

Théorème 1.1.3 (Dai (1995)) *Soit $\overline{X}(\cdot)$ une limite fluide, alors $(\overline{Q}(\cdot), \overline{T}(\cdot))$ est la solution fluide.*

Une conséquence immédiate du théorème ci-dessus est que $\mathfrak{FLM} \subset \mathfrak{FN}$.

Dans ce qui suit, nous abordons une question fondamentale dans la théorie des files d'attente, à savoir pour dériver des conditions qui caractérisent le fait que le réseau a un équilibre unique, dans le sens qu'il existe une mesure de probabilité invariante attractive. Dans ce qui suit nous présentons la définition d'un réseau de files d'attente stable selon Schönlein (2012).

Définition 1.1.3 *Un réseau de files d'attente multiclasse est dit stable si le processus de Markov sous-jacent correspondant X est récurrent positif au sens de Harris.*

Selon la définition de la récurrence positive au sens de Harris, la premiére étape vers une condition suffisante de la stabilité d'un réseau de files d'attente multiclasse est de chercher les petits ensembles fermés. Le lemme suivant montre que Hypothéses 1.1.1.1 fournissent un petit ensemble fermé.

Lemme 1.1.2 (Dai (1995)) *Si les temps des inter-arrivées satisfont les Hypothéses 1.1.1.1, alors pour tout $\kappa > 0$ l'ensemble*

$$A = \{x \in \chi : \ |x| \leq \kappa\}, \tag{1.20}$$

est petit et férmé.

Maintenant, nous énonçons le résultat principal de cette partie, à savoir un critére suffisant pour la stabilité d'un réseau de files d'attente multiclasse en fonction de son modèle de limite fluide associé.

A cette fin, nous introduisons le concept de stabilité pour le modéle de limite fluide selon Schönlein (2012). Cela se fait dans la définition suivante.

Définition 1.1.4 *Un modéle de limite fluide d'une file d'attente sous une discipline donnée est dit stable s'il existe $\tau > 0$ de telle sorte que pour chaque limite fluide $\overline{X}(\cdot) \in \mathfrak{FLM}$ la composante $\overline{Q}(\cdot)$ satisfait $\overline{Q}(t) = 0$ pour tout $t \geq \tau||\overline{Q}(0)||$.*

Théorème 1.1.4 (Dai (1995)) *Etant donnée une discipline de file d'attente fixée. Supposons que les Hypothèses 1.1.1.1 sont satisfaites. Si le modéle de limite fluide est stable, alors le réseau de files d'attente est stable. En particulier, si le réseau fluide associé est stable, alors le réseau de files d'attente est stable.*

En général, il n'est pas facile de travailler avec les limites fluides. Du fait que le réseau fluide associé est un modéle déterministe, alors nous travaillons avec le réseau fluide associé.

Remarque 1.1.3 *L'hypothése 1.1.1.1 (3) apparaît explicitement dans le lemme 2.3.1. Ces conditions imposent des restrictions appropriées sur les distributions des temps d'inter-arrivées. Ainsi, on peut se permettre une distribution générale des temps d'inter-arrivées et montrer directement que la condition (1.20) est satisfaite pour une situation particulière.*

La relation entre les réseaux de files d'attente multiclasses et leurs modèles de limites fluides et les réseaux fluides associés est le contenu de la remarque suivante.

Remarque 1.1.4 *(a) Si le modéle de limite fluide (resp. Le réseau fluide associé) est faiblement instable, c'est à dire que pour chaque trajectoire de l'échantillon $\omega \in G$ il existe $\delta > 0$ qui peux dépendre de ω de telle sorte que $\overline{Q}(\delta) \neq 0$ pour chaque $\overline{Q}(\cdot) \in \mathfrak{FL}(\omega)$ (resp. $\overline{Q}(\cdot) \in \mathfrak{FM}$) avec $\overline{Q}(0) = 0$, alors le réseau de files d'attente est instable dans le sens où presque sûrement, nous avons*

$$\lim_{t \longrightarrow \infty} ||\overline{Q}(t)|| = \infty.$$

(b) Bramson (2008) a prouvé par un contre exemple qu'il existe des réseaux de files d'attente multiclasses stables où le réseau fluide associé est instable. C'est une réciproque à la deuxième notification du théorème 1.1.4.
(c) En outre, un contre exemple par Dai, Hasenbein et Vande Vate (2004) montre que la stabilité d'un réseau de files d'attente multiclasse peut dépendre des distributions des incréments primitifs. En réalité, connaissant les valeurs moyennes des incréments primitifs peut ne pas être suffisant pour conclure la stabilité. Par conséquent, en général le réseau fluide associé n'est pas capable de décrire complétement la stabilité du réseau de files d'attente.
(d) Nous trouvons encore des résultats partiels réciproques dans Meyn (1995), Pukhalskij et Rybko (2004).

Dans ce qui suit nous nous intéressons aux propriétés des réseaux fluides associées aux réseaux de files d'attente sous la discipline de service non-idling. A cette fin, nous considérons l'ensemble des solutions associées aux équations fluides de base (1.13)-(1.18).

Proposition 1.1.2 (Chen (1995)) *(continuité lipschitzienne) Les solutions fluides $(Q(\cdot), T(\cdot))$ sont continues lipschitziennes avec une constante de Lipschitz globale. En particulier, le couple est différentiable presque partout par rapport à la mesure de Lebesgue sur $[0, \infty)$.*

Maintenant, afin de présenter d'autres propriétés, nous introduisons les opérateurs de changement d'échelles d'espace et de temps (l'opérateur de scaling) et l'opérateur de shift (l'opérateur de décalage).

Etant donnée une fonction $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+^K$, $r > 0$, l'opérateur de scaling σ_r est définit par

$$\sigma_r f(t) = \frac{1}{r} f(rt), \tag{1.21}$$

et pour $s \geq 0$ l'opérateur de shift σ_r est définit par

$$\sigma_s f(t) = f(t+s) \tag{1.22}$$

La proposition suivante est une propriété de scaling et de shift des solutions fluides.

Proposition 1.1.3 (Chen (1995)) *Soit $(Q(\cdot), T(\cdot))$ la solution fluide et $Q(0)$ le niveau de fluide initial.*
(1) Pour chaque $r > 0$ le couple $(\sigma_r Q(\cdot), \sigma_r T(\cdot))$ est la solution fluide de niveau fluide initial $\sigma_r Q(0)$.
(2) Pour chaque $s \geq 0$ le couple $(\sigma_s Q(\cdot), \sigma_s T(\cdot) - \sigma_s T(0))$ est la solution fluide de niveau fluide initial $\sigma_s Q(0)$.

La propriété suivante concerne la convergence des processus de niveau fluide.

Proposition 1.1.4 (Ye et Chen (2001)) *(La fermeture pour la topologie uniforme)*
Soit $(Q_n(\cdot), T_n(\cdot))_{n \longrightarrow N}$ une suite de solutions fluides de niveau fluide initial $Q_n(0)$. On suppose que $(Q_n(\cdot), T_n(\cdot))_{n \longrightarrow N}$ converge uniformément sur des ensembles compacts vers $(Q_*(\cdot), T_*(\cdot))$ quand $n \longrightarrow \infty$. Alors $(Q_*(\cdot), T_*(\cdot))$ est une solution fluide de niveau fluide initial $Q_*(0)$.

1.2 Stabilité des réseaux fluides

La méthode d'approximation fluide pour l'analyse de la stabilité des réseaux de files d'attente multi-classes a pour but d'établir les conditions de stabilité pour les réseaux fluides. Ainsi, dans cette partie, nous nous concentrons sur les conditions de stabilité pour les réseaux fluides opérant sous les disciplines de service introduites dans la partie précédente. Nous montrons que pour chaque discipline de service individuelle il existe des différents critéres de stabilité.

Dans un premier temps, nous considérons les réseaux fluides sous les disciplines de services general work-conserving, la seule spécification pour cette classe de réseaux est la propriété non-idling, une station peut prendre un repos dans un seul cas où la file d'attente est vide, voir la condition (**??**). Il sera montré que "la charge de travail nominale de chaque station strictement inférieure à 1" est une condition nécessaire de la stabilité . Notons que pour toutes les disciplines déjà présentées précédemment, cette condition est également nécessaire pour la stabilité du réseau.
De plus, dans cette partie nous présentons les conditions de stabilité suffisantes pour chaque discipline de service étudiée dans la littérature.

Dans les sections suivantes les réseaux fluides sous des disciplines particuliéres sont examinés en détail. La derniére section de ce chapitre fournit une comparaison avec la théorie de la stabilité des systèmes dynamiques modélisés par les équations différentielles.

1.2.1 Généralités

Un réseau fluide comme il a été présenté précédemment est continu déterministe analogue à un réseau de files d'attente multiclasse. Par conséquent les variables décrivant les réseaux fluides sont utilisées. Cependant, l'interprétation est différente. Un réseau fluide se compose de J stations de service et K différentes classes de fluides. Pour chaque classe fluide k la variable α_k est interprétée comme le taux d'écoulement (d'arrivée) des fluides au réseau. Le vecteur correspondant $\alpha_k \in \mathbb{R}_+^K$ est appelé le taux d'arrivée exogène. De même, la variable $\mu_k \in \mathbb{R}_+$ est interprétée comme le taux de départ potentiel des fluides de classe k, et la matrice substochastique $P \in [0,1]^{K\times K}$ est considérée comme la matrice de transition de flux. De plus, il est supposé que $\varrho(P) < 1$, avec ($\varrho(P)$ est le rayon spéctral de la matrice P). Cependant, pour garder l'analogie et la simplification, nous nous référons à $\alpha, \mu,$ et P comme étant le taux d'arrivée, la capacité de service, et la matrice de routage, respectivement. L'application c et la matrice de composition C correspondante sont complétement analogues au cas de réseau de files d'attente.

Le niveau fluide initial du réseau et le processus de niveau fluide sont désignés par $Q(0) \in \mathbb{R}_+^K$ et $Q := \{Q(t) \in \mathbb{R}_+^K, \quad t \geq 0\}$ respectivement. Etant donnés les paramétres α et μ, la structure P et C et le niveau fluide initial $Q(0)$, l'évolution du réseau fluide est déterminée par la discipline de service. Pour un réseau de file d'attente, le processus d'allocation noté par $T := \{T(t) \in \mathbb{R}_+^K, \quad t \geq 0\}$ représente la discipline, où $T_k(t)$ désigne le temps de service accumulé dans la période $[0,t]$ que la station $c(k)$ a alloué pour servir le fluide de classe k. Par conséquent, la quantité $\mu_k T_k(t)$ représente le départ cumulatif des fluides de la classe k Jusqu'à temps t.

Rappelons que les processus sont Lipschitziens et ainsi différentiables presque partout. En outre, l'ensemble des équations dites équations fluides de base définissant le réseau fluide est comme suit

$$Q(t) = Q(0) + \alpha t - (I - P^T)MT(t) \geq 0, \tag{1.23}$$

$$T(\cdot) \quad \text{est} \quad \text{croissant}, \quad \text{avec } T(0) = 0, \tag{1.24}$$

$$I(t) = et - CT(t), \qquad I(\cdot) \quad \text{est} \quad \text{croissant}, \tag{1.25}$$

$$(CQ(t))^T \dot{I}(t) = 0 \quad \text{pour} \quad \text{tout} \quad t \geq 0, \tag{1.26}$$

avec $M = diag(\mu)$ et $I = \{I(t), \quad t \geq 0\}$ est le processus de repos cumulatif. Pour spécifier le réseau sous une discipline de service particulière, nous devons ajouter au moins une autre équation décrivant la discipline. Puisque un réseau fluide est définit par les paraméres α, μ, P, C et la discipline π, alors nous dénotons un réseau fluide par (α, μ, P, C, π). Maintenant, nous rappelons une définition formelle de la stabilité des réseaux fluides selon Schönlein (2012).

Définition 1.2.1 *Un réseau fluide (α, μ, P, C, π) est dit stable s'il existe un $\tau > 0$ de telle sorte que $Q(t) \equiv 0$ pour tout $t \geq \tau \parallel Q(0) \parallel$ et pour tout processus de niveau fluide $Q(\cdot)$.*

Dans le théorème suivant nous introduisons une condition faible de stabilité des réseaux fluides. Il s'avère que la caractérisation suivante de la stabilité est utile pour fournir une comparaison avec la stabilité au sens de Lyapunov pour les systèmes dynamiques.

Pour plus de simplification, nous dénotons par $\Phi(1)$ l'ensemble des processus de niveau fluide avec un niveau initial total égal à 1, c'est à dire $\| Q(0) \| = 1$.

Théorème 1.2.1 (Stoylar (1995)) *Un réseau fluide est stable si et seulement si pour chaque solution fluide $Q(\cdot) \in \Phi(1)$ on a*

$$\inf_{t \geq 0} \| Q(t) \| < \| Q(0) \| = 1. \tag{1.27}$$

Avant de présenter la méthode de Lyapunov pour les réseaux fluides nous considérons le lemme auxiliaire suivant.

Lemme 1.2.1 (Dai (1999)) *Soit $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ une fonction absolument continue et soit $\epsilon > 0$ fixé. Supposons que si $f(t) > 0$ et différentiable au point t, on a $\dot{f}(t) \leq -\epsilon$. Alors, $f(t) = 0$ pour tout $t \geq \epsilon^{-1} f(0)$.*

En se basant sur le dernier résultat, nous rappelons le critére de Lyapunov pour la stabilité des réseaux fluides. Ce dernier est un outil important pour établir des conditions de stabilité sous des disciplines particulières.

Théorème 1.2.2 (Dai (1999)) *On considére le réseau fluide (α, μ, P, C, π). Soit $V : \mathbb{R}_+^K \longrightarrow \mathbb{R}_+$ une fonction localement libschitzienne telle que $V(x) = 0$ si et seulement si $x = 0$. Soit $\epsilon > 0$ et supposons que pour chaque solution fluide $Q(\cdot)$, on a*

$$\dot{V}(Q(t)) := \frac{d}{dt}[V(Q(t))] \leq -\epsilon, \tag{1.28}$$

pour tout $Q(t) \neq 0$ et t est un point régulier pour $s \longrightarrow V(Q(s))$. Alors, le réseau fluide est stable.

Les propriétés de la fonction V, à savoir la positivité et la diminution de chaque processus de niveau fluide avec un taux uniforme, ressemblent aux caractéristiques d'une fonction de Lyapunov pour une équation différentielle ordinaire. Pour cette raison, nous considérons V comme une fonction de Lyapunov. Dans ce qui suit, nous présentons les conditions nécessaires et suffisantes pour la stabilité des réseaux fluides sous des disciplines de service déjà introduites. Pour établir les conditions de stabilité suffisantes il faut trouver des fonctions de Lyapunov appropriées vérifiant le théorème 1.2.2.

1.2.2 Réseaux fluides sous la discipline de service general work-conserving

Nous considérons les propriétés de stabilité des réseaux fluides définies par les équations fluides de base. Comme la seule restriction appropriée est l'équation (1.26). Ces réseaux sont appelés réseaux fuides sous la discipline de service general work-conserving. Par la Proposition 1.1.2, les processus sont différentiables presque partout. Par conséquent, la propriété non-idling (work-conserving) est donnée par

$$\sum_{k \in C(j)} Q_k(t) \dot{I}_j(t) = \left(\sum_{k \in C(j)} Q_k(t) \right) \left(1 - \sum_{k \in C(j)} \dot{T}_k(t) \right) = 0.$$

Le théorème suivant présente l'existence du processus d'allocation pour tout réseau fluide de niveau fluide initial $Q(0)$, sous la discipline de service non-idling. Ce résultat a ètè montrée par Chen (1995).

Théorème 1.2.3 (Chen (1995)) *Pour tout réseau fluide (α, μ, P, C) de niveau fluide initial $Q(0)$, il existe au moins un processus d'allocation $T(\cdot)$ sous la discipline de service non-idling.*

Maintenant, nous présentons un résultat qui montre que la condition de la charge de travail nominale $\rho < e$ est nécessaire pour la stabilité, où $\rho = CM^{-1}(I - P^T)^{-1}\alpha$.

Théorème 1.2.4 (Chen (1995)) *Supposons que le réseau fluide (α, μ, P, C) sous general work-conserving est stable, alors la condition de la charge de travail nominal $\rho < e$ se tient.*

Notons ici que la réciproque de ce dernier résultat est valable si $J = 1$ ou $J = K$. Afin de formuler une condition suffisante pour la stabilité d'un réseau fluide sous les disciplines de services general work-conserving, nous devons introduire les matrices strictement copositives.

Rappelons qu'une matrice symétrique $A \in \mathbb{R}^{K\times K}$ est appelée une matrice strictement copositive si pour chaque $x \in \mathbb{R}_+^K$, on a $x^T Ax \geq 0$ et $x^T Ax = 0$ si et seulement si $x = 0$. En outre, pour $c \in \mathbb{R}$ nous notons par $c^- = \min\{c, 0\}$ et par $c^+ = \max\{c, 0\}$

Théorème 1.2.5 (Chen (1995)) *Considérons un réseau fluide sous la discipline de service non-idling. Supposons que A est une matrice copositive symétrique telle que pour $k = 1, ..., K$*

$$\theta_k = -\left(\sum_{i=1}^{K} \alpha_i a_{ik} - \min_{i\in C(c(k))} h_{ik} - \sum_{j=1, j\neq c(k)}^{J} (\min_{i\in C(j)} h_{ik})^- > 0\right),$$

où $H = M(I - P)A$. Alors le réseau est stable.

Notons que la classe des réseaux fluides sous la discipline de service general work-conserving contient tous les réseaux fluides sous la discipline de service "non-idling" puisque les équations de dynamique d'un réseau fluide sous une certaine discipline "non-idling" sont une spécification des équations de dynamique de réseaux de files d'attente (1.23)-(1.26). Donc, si le réseau fluide sous la discipline de service non-idling est stable alors le réseau est stable sous toutes les disciplines de services non idling. Pour cette raison, la stabilité des réseaux fluides sous les disciplines de services general work-conserving est appelée la stabilité globale.

1.2.3 Réseaux fluides sous des disciplines de service avec priorité

Dans un régime de priorité, les différentes classes de fluides sont servies dans les stations selon un ordre de priorité prédéfini. La priorité est déterminée par une permutation $\pi : \{1, ..., K\} \longrightarrow \{1, ..., K\}$. Etant données les classes de fluide $l, k \in \{1, ..., K\}$ sérvies à la même station $c(l) = c(k)$, les fluides de la classe l sont de priorité plus élevée que les fluides de classe k si $\pi(l) < \pi(k)$. Donc, les fluides de classe k ne sont pas servis tant que le niveau fluide de classe l est supérieur à zéro. Pour chaque fluide de classe

$k \in \{1, ..., K\}$ l'ensemble

$$\Pi_k = \{l : l \in C(c(k)), \pi(l) < \pi(k)\}$$

contient toutes les classes de fluide servies à la même station $c(k)$ et qui ont une priorité plus élevée que les fluides de la classe k. Dans ce qui suit le symbole Π est utilisé pour exprimer la discipline de priorité. Pour décrire les équations de dynamique d'un réseau fluide avec priorité (α, μ, P, C, Π) nous introduisons le processus de capacité inutilisée $Y(t)$, avec $Y_k(t)$ désigne la capacité cumulative restante de $c(k)$ pour servir les classes de fluides qui ont une priorité inférieure à celle de fluide de classe k. Les équations de dynamique peuvent être résumées comme suit

$$Q(t) = Q(0) + \alpha t - (I - P^T)MT(t) \geq 0, \tag{1.29}$$

$$T(\cdot) \text{ est croissante, avec } T(0) = 0, \tag{1.30}$$

$$Y_k(t) = t - \sum_{l \in \Pi_k} T_l(t) \text{ et } Y(\cdot) \text{ est croissante,} \tag{1.31}$$

$$Q_k(t)\dot{Y}_k(t) = 0 \quad \text{pour presque tout } t \geq 0, \quad k \in \{1, ..., K\}. \tag{1.32}$$

Cette représentation est proche des équations de dynamique d'un réseau fluide sous la discipline de service general work-conserving.

Dans ce qui suit nous présentons des conditions pour que le réseau fluide (α, μ, P, C, Π) soit stable. A cette fin, nous donnons une autre description des équations de dynamique plus appropriées pour l'analyse de stabilité.

Soit $\Pi_k^+ = \Pi_k \backslash \{k\}$, et dans le cas où la classe $h(k)$ n'est pas prioritaire à la station $c(k)$ nous avons,

$$h(k) = \arg\max\{\pi(l) : l \in \Pi_k^+\},$$

et $h(k) = 0$ si $\Pi_k^+ = \emptyset$.

Nous considérons aussi la matrice $B \in \mathbb{R}^{K \times K}$ définie par

$$b_{lk} = \begin{cases} 1, & h(k) = l \\ 0, & \text{sinon} \end{cases} \tag{1.33}$$

Soit e^{Π} définie par

$$e_k^{\Pi} = \begin{cases} 1, & Pi_k^+ = \emptyset \\ 0, & \text{sinon} \end{cases} \tag{1.34}$$

Avec cette notation le processus d'allocation $T(\cdot)$ peut être exprimé par

$$T(t) = -(I - B)Y(t) + e^{\Pi}t.$$

De plus, en prenant $\theta = \alpha - (I - P^T)Me^{\Pi}$ et $R = (I - P^T)M(I - B)$. L'équation d'équilibre (1.29) sera comme suit

$$Q(t) = Q(0) + \theta t + RY(t) \geq 0. \tag{1.35}$$

Encore, la condition de priorité peut être exprimée aussi par la condition suivante

$$\forall l,k \in \{1,...,K\},\ \ \forall t \geq 0 \ \text{(régulier)} :\ \ c(l) = c(k), \pi(l) < \pi(k) \Longrightarrow \dot{Y}_l(t) \geq \dot{Y}_k(t). \tag{1.36}$$

Pour mieux comprendre la discipline de priorité, nous considérons les partitions (a,b) de l'ensemble des classes fluides $\{1,...,K\}$ prennant en compte la discipline de priorité, dans le sens où, si la classe fluide l est dans a et la classe fluide k est servie à la station $c(l)$, de plus la classe k a une priorité plus élevée que l, alors la classe k est dans a. Pour plus de details, nous introduisons la définition suivante donnée par Schönlein (2012).

Définition 1.2.2 *Une partition (a,b) de $\{1,...,K\}$ est dite hiérarchique par rapport à π si $l \in a$ et $k \in \Pi_l^+$, alors $k \in a$. L'ensemble de toutes les partitions hiérarchiques de $\{1,...,K\}$ est noté par $\mathcal{H}$.*

Soit $(a,b) \in \mathcal{H}$ une partition hiérarchique, alors l'équation fluide d'équilibre (1.35) sous forme matricielle se donnée comme suit

$$\begin{pmatrix} Q_a(t) \\ Q_b(t) \end{pmatrix} = \begin{pmatrix} Q_a(0) \\ Q_b(0) \end{pmatrix} + \begin{pmatrix} \theta_a \\ \theta_b \end{pmatrix} t + \begin{pmatrix} R_a & R_{ab} \\ R_{ba} & R_b \end{pmatrix} \begin{pmatrix} Y_a(t) \\ Y_b(t) \end{pmatrix}. \tag{1.37}$$

De plus, pour $(a,b) \in \mathcal{H}$ avec $b \neq \emptyset$ nous définissons

$$S(b) = \{u \geq 0, \theta_a + R_b u = 0 \ \text{ et } \ u \leq e\}$$

Théorème 1.2.6 (Chen et Zhang (2000)) *Nous considérons le réseau fluide (α, μ, P, C, Π). Nous assumons que $\rho < e$ et qu'il existe $\varepsilon > 0$, et un vecteur $h \in \mathbb{R}_+^K$ de telle sorte que pour chaque partition $(a,b) \in \mathcal{H}$ la condition*

$$h_a^T(\theta_a + R_{ab}x_b) < -\varepsilon, \tag{1.38}$$

se tient pour $x_b \in S(b)$ si $b \neq \emptyset$, et $x_b = 0$ si $b = \emptyset$. De plus si $S(b) = \emptyset$ la condition (1.38) se tient également. Alors, le réseau fluide est stable.

La condition suffisante du théorème précédent peut être affaiblie, comme on peut le voir ci-après. Afin de décrire une condition suffisante plus générale, quelques notations seront considérées. Pour $b \subseteq \{1,...,K\}$ l'ensemble

$$\Pi(b) = \{k \in \{1,...,K\} : l \in b \text{ et } c(k) = c(l) \Longrightarrow \pi(k) < \pi(l)\},$$

définit l'ensemble des classes de fluide prioritaires dans chaque station qui sert au moins une classe de fluide dans "b". Pour chaque partition $(a,b) \in \mathcal{H}$, l'ensemble

$$S(a,b) = \{q \in \mathbb{R}^K : q \geq 0, q_a = 0 \text{ et } q_{\Pi(b)} > 0\},$$

définit l'état de fluide. En outre, nous considérons la condition

$$0 \leq x \leq e \quad \text{et} \quad \theta_a + R_a x = 0, \tag{1.39}$$

et nous définissons l'ensemble des taux des flux réguliers par

$$F(a,b) = \{d = (0 \ d_b)^T \in \mathbb{R}^K : d_b = \theta_b + R_{ba} y_a \quad \text{où} \quad y_a \ \text{ satisfait } \ (1.39)\}.$$

Autrement dit, les équations de dynamique impliquent que pour chaque partition $(a,b) \in \mathcal{H}$, si t est un point régulier et $Q(t) \in S(a,b)$, alors nous avons $\dot{Q}(t) \in F(a,b)$.
En fait, $F(a,b)$ est exactement l'ensemble de toutes ces dérivées.

Théorème 1.2.7 (Chen et Ye (2002)) *On suppose qu'il existe $\varepsilon > 0$, un entier $N \geq 1$ et N vecteurs positifs de dimensions K, $h_1, ..., h_N$ de telle sorte que les conditions suivantes sont satisfaites.*

(a) Associé à chaque partition $(a,b) \in \mathcal{H}$ avec $b \neq \emptyset$ et $F(a,b) \neq 0$ il existe un ensemble d'indice $I(a,b) \subseteq \{1, ..., N\}$ tel que pour tout $i \in I(a,b)$

$$\sup_{d \in F(a,b)} h_i^T d \leq -\varepsilon. \tag{1.40}$$

(b) Pour chaque partition $(a,b) \in H$ avec $b \neq \emptyset$ et $F(a,b) \neq \emptyset$ et pour chaque $j \notin I(a,b)$, il existe $i \in I(a,b)$ de telle sorte que

$$(h_j)_b \leq (h_i)_b. \tag{1.41}$$

Alors

$$V(x) = \max_{1 \leq i \leq N} h_i^T x$$

est une fonction de Lyapunov linéaire par morceaux. En particulier, le réseau fluide (α, μ, P, C, Π) est stable.

1.2.4 Réseaux fluides sous la discipline de service FIFO

Pour les réseaux fluides sous des disciplines de service FIFO les clients sont servis dans l'ordre premier arrivé premier sorti. Pour décrire l'évolution des classes de fluides k, nous devons considérer la charge de travail immédiate donnée par $W(t) = CM^{-1}Q(t)$. Toute fluide arrivée aprés le temps t a une priorité inférieure sous la discipline de service FIFO. Ainsi, les fluides qui arrivent au temps t sont servies au temps $t + W_j(t)$ avec $W_j(t)$ appelée la charge de travail de la station $j = c(k)$. Toutes les arrivées jusqu'au temps t sont données par

$$A(t) = \alpha t + P^T MT(t).$$

Pour chaque classe $k \in \{1, ..., K\}$ le régime sous la discipline de service FIFO peut être représenté par la relation suivante

$$T_k(t + W_j(t)) = m_k(Q_k(0) + A_k(t)). \tag{1.42}$$

Avec $m_k = \mu_k^{-1}$. Ainsi, les équations décrivant un réseau fluide sous FIFO peuvent être récapitulées comme suit

$$Q(t) = Q(0) + \alpha t - (I - P^T)MT(t) \geq 0, \tag{1.43}$$

$$T(\cdot) \text{ est croissant, avec } T(0) = 0, \tag{1.44}$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ est croissant}, \tag{1.45}$$

$$(CQ(t))^T \dot{I}(t) = 0 \text{ pour presque tout } t \geq 0, \tag{1.46}$$

$$T_k(t + W_j(t)) = m_k(Q_k(0) + A_k(t)) \text{ pour tout } k \in \{1, ..., K\}. \tag{1.47}$$

Notons qu'un réseau fluide sous la discipline de service FIFO n'est pas complétement déterminé par le niveau fluide initial $Q(0)$, c'est parce qu'il faut préciser comment le niveau fluide initial est servi dans $[0, W_j(0)]$. Ainsi, les charges initiales pour chaque classe $k \in \{1, ..., K\}$ sont données par $Q(0)$ et

$$\{T_k(s) : s \in [0, W_j(0)]\}. \tag{1.48}$$

Remarque 1.2.1 *Bramson (1994) et Seidman (1994) ont montré un résultat inattendu, ils ont montré que les réseaux fluides sous la discipline de service FIFO satisfaisant la condition de charge de travail nominale ne sont pas stables en général. En outre, Bramson (1996) a également montré que les réseaux fluides sous la discipline de service FIFO de type Kelly (les capacités de service des classes présentes à la même station sont égales) sont stables si et seulement si la condition de charge de travail nominale est satisfaite.*

Pour donner des conditions de stabilité suffisantes, obtenues par Chen et Zhang (1997), nous supposons que le vecteur des taux d'arrivées effectifs est donné par $\lambda = (I - P^T)^{-1}\alpha$ avec $\Lambda = diag(\lambda)$ et le rayon spectral est désigné par ϱ.

Théorème 1.2.8 (Chen et Zhang (1997)) *Supposons que le réseau fluide sous la discipline de service FIFO satisfait $\rho < e$ et une des conditions suivantes*
(i) $\varrho(P^T + (I - P^T)\Lambda C^T CM) < 1$.

(ii) $\varrho(CMP^T(I - P^T)\Lambda C^T) < 1$.
Alors, le réseau fluide sous la discipline de service FIFO est stable.

Ceci indique que la discipline de service FIFO est spéciale parmi les disciplines de services considérées dans cette thèse.

1.2.5 Théorie de la stabilité des équations différentielles

La notion de la stabilité présentée dans la définition (1.2.1) peut également être interprétée comme le processus de niveau fluide zéro $Q_0(\cdot) \equiv 0$, étant l'unique point fixe, attractif et stable de l'opérateur de shift $\delta_\tau Q(\cdot) = Q(\tau + \cdot)$ défini sur l'ensemble des processus de niveau fluide. Pour voir cela, supposons

que $Q_*(\cdot)$ est un autre point fixe. Alors, pour tout $t \geq 0$ nous avons

$$Q_*(t) - Q_0(t) = \delta_\tau Q_*(t) = 0,$$

où la dernière égalité est valable puisque le réseau est stable.

Maintenant, nous rappelons brièvement la définition de la stabilité et les fonctions de Lyapunov de la théorie de systèmes dynamiques. Pour une description détaillée, nous pouvons voir Bacciotti et Rosier (2005), Hinrichsen et Pritchard (2005).

Considérons l'équation différentielle ordinaire

$$\dot{x}(t) = f(x(t)), \quad x \in \mathbb{R}^n, \quad t \in [0, \infty), \tag{1.49}$$

avec la condition initial $x(0) = x_0$. Et f une fonction continue, avec $f(0) = 0$ (l'origine est un point d'équilibre). L'origine serait dit globalement asymptotiquement stable au sens de Lyapunov si

Stabilité :

Pour tout $\varepsilon > 0$ il existe $\delta > 0$ de telle sorte que $||x_0|| < \delta \Rightarrow ||x(t)|| < \varepsilon$ pour tout $t \geq 0$ et pour tout $x(\cdot)$ avec $x_0 = x(0)$.

Attractivité :

Il existe $\eta > 0$ de telle sorte que pour tout $||x_0|| < \eta \Rightarrow$ pour tout $x(\cdot)$ avec $x_0 = x(0)$, $\lim\limits_{t \longrightarrow \infty} ||x(t)|| = 0$ et η peut être prise aussi grande que souhaitée.

Une autre façon de décrire la stabilité est d'utiliser la fonction de Lyapunov et les fonctions de comparaison.

Une fonction $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ est dite de classe $\mathcal{K}$ si elle est continue et strictement croissante et satisfait $f(0) = 0$. L'application $V : \mathbb{R}^n \longrightarrow \mathbb{R}$ à valeur rélle est appelée fonction de Lyapunov stricte globale pour (1.49) si

(i) Elle est définie positive et propre (radialement non bornée) c'est à dire ; il existe des fonctions a, b de classe $\mathcal{K}$ de telle sorte que pour tout $x \in \mathbb{R}^n$ nous avons

$$a(||x||) \leq V(x) \leq b(||x||). \tag{1.50}$$

(ii) Il existe des fonctions w de classe $\mathcal{K}$ tels que pour toute solution $x(\cdot)$ et chaque intervalle $I \subset [0, \infty)$ nous avons

$$V(x(t)) - V(x(s)) \leq -\int_s^t w(||x(r)||)dr, \tag{1.51}$$

pour chaque $s < t \in I$ à condition que $x(\cdot)$ soit dans I.

Remarque 1.2.2 *Il est bien connu que l'origine est globalement asymptotiquement stable si et seulement s'il existe une fonction de Lyapunov stricte globale, voir Bacciotti et Rosier (2005). De ce point de vue, la définition de la stabilité des réseaux fluides semble s'écarter de la stabilité asymptotique au sens de Lyapunov. Cependant, le lemme suivant prouve que les définitions de la stabilité sont en réalité équivalentes.*

Lemme 1.2.2 (Schönlein (2012)) *Un réseau fluide est stable si et seulement si le processus de niveau fluide zéro est globalement asymptotiquement stable au sens de Lyapunov.*

Compte tenu du théorème (1.2.2) et la définition des fonctions de Lyapunov dans la théorie des systèmes dynamiques, nous fournissons maintenant une définition précise d'une fonction de Lyapunov pour la classe des réeaux fluides selon Schönlein (2012).

Définition 1.2.3 *étant donné un réseau fluide, la fonction* $V : \mathbb{R}_+^k \longrightarrow \mathbb{R}$ *définie positive et propre est appelée fonction de Lyapunov s'il existe une fonction* $w \in \mathcal{K}$ *de telle sorte que*

$$V(Q(t)) - V(Q(s)) \leq -\int_s^t w(||Q(r)||)dr, \tag{1.52}$$

pour tout $0 \leq s \leq t$ *et pour tout processus de niveau fluide* $Q(\cdot)$.

La fonction V définie dans le théorème (1.2.2) est supposée localement Lipschitzienne, définie positive, et satisfait la condition de décroissance uniforme.

$$\dot{V}(Q(t)) \leq -\varepsilon \tag{1.53}$$

si $Q(t) \neq 0$ et t régulier pour toute application $s \longrightarrow V(Q(s))$. Nous notons que si V est localement Lipschitzienne, la condition de décroissance de la définition (1.2.3) sera donnée comme suit

$$\dot{V}(Q(t)) \leq -w(||Q(t)||), \tag{1.54}$$

où la décroissance dépend du niveau de fluide à l'instant actuel, plutôt que d'être uniforme. En tout cas la fonction de Lyapunov V dans le théorème 1.2.2 n'est pas forcément propre .

Dans la théorie de Lyapunov pour les systémes dynamiques cette condition est essentielle pour qu'un équilibre soit stable car cela renforce le caractère fermé des ensembles de sous-niveau de la fonction de Lyapunov.

Cela garantit que les trajectoires, le long desquelles la fonction de Lyapunov décroit, convergeront vers l'équilibre et ne divergeront pas. Cependant, en raison de la propriété de scaling, ce comportement est négligeable pour les réseaux fluides car la décroissance de la fonction de Lyapunov V est uniforme, voir le théorème 1.2.2.

1.3 Contribution de la thèse

Dans cette thèse nous étudions la stabilité de différents réseaux de files d'attente fluides multiclasses en utilisant une approche qui vise à réduire le problème de la stabilité de tels modèles à l'analyse de la stabilité de réseaux pûrement déterministes, appelés réseaux fluides associés. A cette fin plusieurs résultats ont été établis :

Résultat 1 : Stability condition of a priority queueing system

Dans ce travail nous établissons la condition de stabilité globale d'un modèle de réseaux de files d'attente multiclasses composé de N unités ($N \geq 3$) et N^2 classes de clients, où à chaque unité un

serveur peut servir N classes de clients. Un client vient de l'éxtérieur au réseau avec un taux α par unité de temps, demandant des services de moyennes m_i de chaque unité, passant d'une unité à une autre alignant toutes les files d'attente du système. Et après avoir été traité par les N unités N^2 fois, il quitte définitivement le réseau. La discipline de service sous l'étude est la suivante : si l'ordre de la station i, $i = \overline{1,N}$ est impair, la classe k appartenant à cette station est prioritaire par rapport la classe d'ordre $k-N$, et si l'ordre de la station i, $i = \overline{1,N}$ est pair, la classe k qui appartient à cette station est prioritaire par rapport la classe $k+N$. Chaque fluide résisdant dans la file k, $k = 1, 2, ..., N^2$ est dite fluide de classe k.

En utilisant l'approache de la fonction de Lyapunov linéaire par morceaux, nous déterminons la région de stabilté globale de tel réseau qui est très difficile sous certaines disciplines de services, même si l'intensité du trafic à chaque unité du réseau est inférieure à 1.

En effet, nous montrons qu'avec les conditions de trafic habituelles

$$\rho = \left(\alpha \sum_{k:\sigma(k)=i} m_k < 1 \right), \text{ pour } i = 1, 2, ..., N,$$

la condition additionnelle suivante :

$$\alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} \leq 1,$$

est suffisante pour assurer la stabilité globale.

Résultat 2 : Stabilizing priority fluid queueing network model

Dans ce deuxième travail nous considérons certaines grandes classes de réseaux de files d'attente fluides multiclasses sous des disciplines de service avec priorité. Spécifiquement, nous établissons la condition de stabilité de certains réseaux fluides hétérogènes sous la discipline de service avec priorité. Les réseaux sous l'étude sont composés de N stations et $2N$ classes de clients, où chaque station peut servir plus d'une classe de client avec un service de priorité différentié, et chaque client peut exiger un service séquentiel par plus d'une station de service. Donc, dans notre cas, la performance du réseau est améliorée même lorsque plusieurs charges de travail sont admises pour le service. Pour stabiliser nos réseaux, un certain nombre de stations devrait être ajouté, ceux-ci agissent plus tard en tant que régulateurs pour les systèmes, l'ajout de ces stations n'est pas aléatoire, il dépend essentiellement de priorité supérieure et inférieure de classes de clients et sur le nombre de stations du rśeaux. L'approche du modèle fluide est utilisée pour prouver la stabilité.

1. En premier lieu, nous avons considéré un réseau de files d'attente multiclasse composé de N stations et $2N$ classes de clients, le processus d'arrivée de la classe de clients k; $k = \overline{1,2N}$ arrivant au système suit une loi de Poisson avec des taux d'arrivée λ_1 et λ_{N+1} (≥ 0) à la station 1 et $N+1$ respectivement, le temps de service pour chaque classe de client k est distribué exponentiellement avec un temps moyen de service $m_k > 0$. Nous supposons aussi que tous les temps d'inter-arrivées est les temps de services sont

indépendants, et que chaque classe d'ordre pair à la station $i = \overline{1, N}$ à une priorité supérieure.

Nous modifions notre réseau de telle sorte que si le nombre de station est d'ordre pair (respectivement. impaire) nous ajoutons N (resp. $N-1$) régulateurs.

Nous montrons que si les conditions de stabilité habituelles sont satisfaites, la condition suffisante pour la stabilité établie dans Chen and Zhang (2000) n'est pas satisfaite.

De plus, si

$$\lambda_{k_1} > (1 - m_{k_1'}/m_{k_1''})/m_{k_1}, \tag{1.55}$$

alors le processus $Q(\cdot)$ est récurrent positif.

Avec

$*$ $Q(\cdot)$ est le processus de la longueur de la file d'attente, λ_{k_1} (resp. m_{k_1}) est le taux d'arrivée exogène (resp. le temps moyen de service) de la classe fluide de priorité supérieure des stations additionnelles $i = \overline{N+1, 2N}$, (N : pair) (resp. $i = \overline{N+1, 2N-1}$, (N : impair)), tel que $k_1 = \overline{3N+1, 4N}$, (N : pair) (resp. $k_1 = \overline{3N, 4N-2}$, (N : impair)).

$*$ $m_{k_1'}$ est le temps moyen de service de la classe fluide de priorité inférieure des stations additionnelles $i = \overline{N+1, 2N}$, (N : pair) (resp. $i = \overline{N+1, 2N-1}$, (N : impair)).

$$m_{k_1'} = \begin{cases} m_{k_1-N}, & k_1 = \overline{3N+1, 4N}, \ \text{(N : pair)}, \\ m_{k_1-(N-1)}, & k_1 = \overline{3N, 4N-2}, \ \text{(N : impair)}. \end{cases}$$

$*$ $m_{k_1''}$ est le temps moyen de service de la classe fluide de priorité supérieure du réseau original.

$$m_{k_1''} = \begin{cases} m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N}{2}}, \ j_1 = \overline{1, \frac{N}{2}} \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+2}{2}, k_1 - N}, \ j_1 = \overline{2, N} \end{cases} \quad \text{(N : pair)}, \\ \begin{cases} m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N-1}{2}}, \ j_1 = \overline{1, \frac{N-1}{2}} \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+1}{2}, k_1 - (N-1)}, \ j_1 = \overline{4, N+1} \end{cases} \quad \text{(N : impair)}.$$

Où pour chaque k_1' correspond k_1 et j_1, (j_1 est un nombre pair).

2. En second lieu nous étudions la stabilité du même réseau déja présenté, mais cette fois-ci nous supposons que la priorité supérieure est consacrée aux classes $N, \overline{N+2, 2N}$.

Nous modifions notre réseau en ajoutant N régulateurs, à comparer avec le réseau original, nous avons N stations additionnelles, à savoir la station $N+1$,.....,station $2N$, telle que la classe $3N+1$ à la priorité supérieure dans la station $N+1$, et les classes $\overline{3N+2, \ 4N}$ sont de priorité supérieure aux stations $\overline{N+2, 2N}$.

Nous montrons que si les conditions de stabilité habituelles sont satisfaites, la condition suffisante pour la stabilité établie dans Chen and Zhang (2000) n'est pas satisfaite.

Si

$$\lambda_{3N+1} > (1 - m_{2N+1}/m_N)/m_{3N+1}, \;\; \lambda_{k_2} > (1 - m_{k_2'}/m_{k_2''})/m_{k_2} \tag{1.56}$$

$k_2'' = \overline{N, 2N}$, $k_2' = \overline{2N+1, 3N}$, $k_2 = \overline{3N+1, 4N}$, où pour chaque k_2 on correspond k_2' et k_2''.

Alors, le processus $Q(\cdot)$ est récurrent positif.

Résultat 3. A Note on an $M/M/s$ queueing system with two reconnect and two redial orbits

Ce dernier résultat présente l'analyse d'un centre d'appel modélisé par un modèle de files d'attente $M/M/s$ dont deux flux exogènes de différents types de clients arrivent au système. Les flux de clients rentrent au système suivant un processus de Poisson indépendants, un client de type i, $i = 1, 2$ est géré par un serveur disponible, sinon il attend dans une file d'attente de taille infinie. Les clients sont traités selon la discipline de service FIFO. Le temps de service requis par chaque client est indépendant de son type.

Un client de type i qui n'est pas servi devient impatient et peut abandonner le système après une certaine période de temps exponentiellement distribuée ψ, telle que $\mathbb{E}(\psi) = \frac{1}{\vartheta} < \infty$, où ϑ est le taux d'abandon. Le client abandonné soit quitte définitivement le système avec une probabilité α et il est considéré comme client perdu ou décide de rester avec une probabilité $(1 - \alpha)$. Donc, si le client décide de ne pas quitter le système, il se dirige vers l'une des orbites réservées aux clients impatients dite "redial orbit". Le choix est aléatoire et ne dépend pas du seuil des orbites ou du type de client. Le client résidant dans l'une des orbites tente encore une fois de recevoir un service avec une probabilité $(1-\alpha)\alpha_i$, $(i = 1, 2)$ $(\alpha_1 + \alpha_2 = 1)$, et recompose après une période de temps exponentiellement distribuée ω_i, (i=1,2) avec $\mathbb{E}(\omega_i) = \gamma_i < \infty$. Nous assumons que le temps de service d'un client de type i suit une distribution exponentielle de moyenne $\frac{1}{\mu}$, les temps de service sont indépendants. Après que le client soit servi, il peut quitter le système avec une probabilité β ou revenir avec une probabilité $(1 - \beta)$. Le client de type i se dirige vers l'une des orbites de type i reservée aux clients revenant de l'extérieur demandant d'autres services dite "reconnect orbit", le client revient avec une probabilité $(1 - \beta)\beta_i$, où $(\beta_1 + \beta_2 = 1)$, et il se connecte au serveur après une période de temps exponentiellement distribuée φ_i, avec $\mathbb{E}(\varphi_i) = \theta_i < \infty$.

Dans ce travail nous utilisons un modèle fluide pour dériver des approximations de premier ordre pour le nombre de clients dans les orbites dans le trafic lourd. La limite fluide d'un tel modèle est la solution unique d'un système à trois équations différentielles.

Le processus fluide renormalisé (scaled) de notre modèle est définit ainsi

$$\bar{Z}^{(n)}(t) = (\bar{Z}_Q^{(n)}(t), \bar{Z}_R^{(n)}(t), \bar{Z}_O^{(n)}(t))^T,$$

avec

$$\bar{Z}_Q^{(n)}(t) = \frac{Z_Q^{(n)}(t)}{n}, \quad \bar{Z}_R^{(n)}(t) = \frac{Z_R^{(n)}(t)}{n}, \quad \bar{Z}_O^{(n)}(t) = \frac{Z_O^{(n)}(t)}{n}.$$

Où $Z_Q(t)$ est le nombre de clients dans la file d'attente plus le nombre de clients en service au temps t,

$Z_R(t)$ est le nombre de clients dans les orbites 1 et 2 réservées aux clients impatients, et $Z_O(t)$ représente le nombre de clients dans les orbites 1 et 2 consacrées aux clients revenant de l'extérieur.

Notons que si $\bar{z}^{(n)}(\cdot) \to^{\mathcal{L}} z(\cdot)$, alors $z(\cdot)$ est appelée limite fluide du modèle stochastique original.

Etant données les valeurs déterminites suivantes $(z_Q(0), z_R(0), z_O(0))$, nous supposons que

$$(\bar{Z}_Q^{(n)}(t), \bar{Z}_R^{(n)}(t), \bar{Z}_O^{(n)}(t)) \to^{\mathcal{L}}_{n\to\infty} (z_Q(0), z_R(0), z_O(0)),$$

Alors, la limite fluide du modèle stochastique original est l'unique solution du système d'équations suivant.

$$\begin{aligned} Z_Q(t) = Z_Q(0) + (\lambda_1+\lambda_2)t + (\gamma_1+\gamma_2)\int_0^t Z_R(u)du + (\theta_1+\theta_2)\int_0^t Z_O(u)du - \\ \mu\int_0^t \min\{s, Z_Q(u)\}du - \vartheta\int_0^t (Z_Q(u)-s)^+ du \end{aligned} \tag{1.57}$$

$$Z_R(t) = Z_R(0) + (1-\alpha)\vartheta\int_0^t (Z_Q(u)-s)^+ du - (\gamma_1+\gamma_2)\int_0^t Z_R(u)du \tag{1.58}$$

$$Z_O(t) = Z_O(0) + (1-\beta)\mu\int_0^t \min\{s, Z_Q(u)\}du - (\theta_1+\theta_2)\int_0^t Z_O(u)du \tag{1.59}$$

Maintenant, afin de dériver la limite fluide en stationnarité, c'est-à-dire développer des conditions sous laquelle $z(t)$ est constante.

Nous différencions les équations (4.27)-(4.29),

nous obtenons

$$\lambda_1+\lambda_2 = \mu\min\{s, z_Q(\infty)\}) + \vartheta(z_Q(\infty)-s)^+ - (\gamma_1+\gamma_2)z_R(\infty) - (\theta_1+\theta_2)z_O(\infty). \tag{1.60}$$

$$(\gamma_1+\gamma_2)z_R(\infty) = (1-\alpha)\vartheta(z_Q(\infty)-s)^+. \tag{1.61}$$

$$(\theta_1+\theta_2)z_O(\infty) = (1-\beta)\mu\ \min\{s, z_Q(\infty)\}. \tag{1.62}$$

Où $z_Q(\infty) = \lim_{t\longrightarrow\infty} z_Q(t)$, $z_R(\infty) = \lim_{t\longrightarrow\infty} z_R(t)$, $z_O(\infty) = \lim_{t\longrightarrow\infty} z_O(t)$.

Et les équations (4.30)-(4.32) peuvent être résolues par rapport à $z_Q(\infty)$, $z_R(\infty)$ et $z_O(\infty)$, ce qui donne

$$z_Q(\infty) = \begin{cases} \frac{\lambda_1+\lambda_2}{\mu\beta} & \text{si } \frac{\lambda_1+\lambda_2}{s\mu} < \beta. \\ \frac{\lambda_1+\lambda_2-\beta\mu s}{\vartheta\alpha} + s & \text{sinon.} \end{cases} \tag{1.63}$$

$$z_R(\infty) = \begin{cases} 0 & \text{si } \frac{\lambda_1+\lambda_2}{s\mu} < \beta. \\ \frac{(1-\alpha)\vartheta(z_Q(\infty)-s)}{\gamma_1+\gamma_2} & \text{sinon.} \end{cases} \tag{1.64}$$

$$z_O(\infty) = \begin{cases} \frac{(1-\beta)\mu z_Q(\infty)}{\theta_1+\theta_2} & \text{si } \frac{\lambda_1+\lambda_2}{s\mu} < \beta. \\ \frac{(1-\beta)\mu s}{\theta_1+\theta_2} & \text{sinon.} \end{cases} \tag{1.65}$$

1.4 Plan de la thèse

Notre étude porte essentiellement sur l'analyse des réseaux multiclasses fluides.

La thèse est composée de quatres chapitres :

Le premier chapitre est un chapitre introductif, il présente un travail de synthèse, avec laquelle la thèse est entamée, dans cette partie de thèse une description détaillée des réseaux de files d'attente multiclasses, du processus de Markov sous-jacent, et des équations de dynamique de réseaux de files d'attente est présentée, de plus la notion de réseau fluide est introduite. Dans ce chapitre les critères de stabilité des réseaux fluides sous diverses disciplines de service ont été étudiées et cela en se basant sur quelques résultats existants. A la fin de ce chapitre une comparaison avec la théorie de Lyapunov pour les systèmes dynamiques modélisés par des équations différentielles ordinaires est présentée.

Le deuxième chapitre est consacré à l'étude de la stabilité d'un système de files d'attente multiclasses avec priorité, dans ce travail nous établissons une condition suffisante pour la stabilité d'un réseau de files d'attente composé de N unités, $N \geq 3$ et N^2 classes de clients (N classes à chaque stations), les clients sont servis selon une discipline de service prioritaire. En utilisant la fonction de Lyapunov linéaire par morceaux, la région de stabilité globale est déterminée. Ainsi ce résultat a fait l'objet d'une publication internationale dans **Mathematical Sciences And Applications E-Notes Volume 1 No. 2 pp. 165-172 (2013) MSAEN**.

Le troisième chapitre est consacré à l'étude de certaines grandes classes de réseaux de files d'attente fluides multiclasses sous des disciplines de service avec priorité. Plus précisement, une condition de stabilité de certains réseaux fluides avec priorité composés de N stations et $2N$ classes de clients est établie. Tels réseaux sont stabilisés par l'ajout d'un certain nombre de stations en se basant sur la priorité inférieure et supérieure de leurs classes de clients et du nombre de stations. L'approche du modèle fluide est utilisée pour l'analyse de la stabilité. Ce travail a fait l'objet de publication dans **Acta Univ. Sapientiae, Mathematica, 6, 2 (2014) 146-161.**

Le quatrième chapitre est dédié à l'analyse d'un système de centre d'appel modélisé par un système de files d'attente $M/M/s$; composé d'une file d'attente de taille infinie, deux type de flux de clients indépendants suivant le processus de Poisson, s serveur, deux orbites réservées aux flux de clients impatients dites "redial orbit" et deux orbites céservées aux clients revenant de l'éxterieur appelée "reconnect orbit". Dans cette étude, un modèle fluide est utilisé pour calculer une approximation de premier ordre pour le nombre de clients dans les orbites du système. Ce travail a été ponctué d'une publication parue dans **Appl. Appl. Math. Vol. 10, Issue 1 (June 2015), pp. 1 - 12.**

Bibliographie

A. Bacciotti and L. Rosier. (2005). Liapunov functions and stability in control theory. 2nd ed. Communications and Control Engineering. Springer, Berlin Heidelberg.

D. Bertsimas, I. Ch. Paschalidis, and J. N. Tsitsiklis. (1994). Optimization of multiclass queueing networks : polyhedral and nonlinear characterizations of achievable performance. Annals of Applied Probability 4 : 43-75.

M. Bramson. (1994). Instability of FIFO queueing networks. Ann. Appl. Probab., 4(2) :414-431.

M. Bramson. (1996). Convergence to equilibria for fluid models of FIFO queueing networks. Queueing Syst., 22(1-2) : 5-45.

M. Bramson. (1996). Convergence to equilibria for fluid models of head-ofthe- line proportional processor sharing queueing networks. Queueing Syst., 23(1-4) : 1-26.

M. Bramson. (2008). Stability of queueing networks. Lecture Notes in Mathematics 1950, Springer, Berlin Heidelberg.

H. Chen. (1995). Fluid approximations and stability of multiclass queueing networks : Work-conserving disciplines. Ann. Appl. Probab., 5(3) : 637-665.

H. Chen. (1995). Fluid approximations and stability of multiclass queueing networks : work-conserving disciplines. Annals of Applied Probability 5 : 637-665.

H. Chen and H. Ye. (2002). Piecewise linear Lyapunov function for the stability of multiclass priority queueing networks. IEEE Trans. Autom. Control, 47(4) : 564-575.

H. Chen and H. Zhang. (1997). Stability of multiclass queueing networks under FIFO service discipline. Math. Oper. Res., 22(3) : 691-725.

H. Chen and H. Zhang. (2000). Stability of multiclass queueing networks under priority service disciplines. Oper. Res., 48(1) : 26-37.

J. G. Dai. (1995). On positive Harris recurrence of multiclass queueing networks : a unified approach via

fluid limit models. Annals of Applied Probability 5 : 49-77.

J. Dai. (1996). A fluid limit model criterion for instability of multiclass queueing networks. Ann. Appl. Probab., 6(3) : 751-757.

J. Dai. (1999). Stability of fluid and stochastic processing networks. Ma- PhySto. Miscellanea. 9. Aarhus : Univ. of Aarhus.

J. Dai. (1995). Stability of open multiclass queueing networks via fluid models. In Kelly, F. and Williams, R. (Eds.), Stochastic Networks. Volume 71 of IMA Volumes in Mathematics and Its Applications, pp. 71-90, Springer New York.

J. Dai, J. J. Hasenbein, and B. Kim. (2007). Stability of join-the-shortestqueue networks. Queueing Syst., 57(4) : 129-145.

J. Dai, J. J. Hasenbein, and J. H. (2004). Vande Vate. Stability and instability of a two-station queueing network. Ann. Appl. Probab., 14(1) : 326-377.

L. C. Evans. (1998). Partial differential equations. Graduate Studies in Mathematics 19. American Mathematical Society, Providence, Rhode Island.

D. Hinrichsen and A. J. (2005). Pritchard. Mathematical systems theory. I. Modelling, state space analysis, stability and robustness. Texts in Applied Mathematics 48. Springer, Berlin Heidelberg.

H. Kaspi and A. Mandelbaum. (1992). Regenerative closed queueing networks. Stochastics Stochastics Rep., 39(4) : 239-258.

S. Kumar and P. R. Kumar. (1994). Performance bounds for queueing networks and scheduling policies. IEEE Transactions on Automatic Control 39 : 1600-1611.

P. R. Kumar and S. P. Meyn. (1995). Stability of queueing networks and scheduling policies. IEEE Transactions on Automatic Control 40 : 251-260.

P. R. Kumar and S. P. Meyn. (1996). Duality and linear programs for stability and performance analysis of queuing networks and scheduling policies. IEEE Transactions on Automatic Control 41 : 4-17.

P. Kumar and T. I. Seidman. (1990). Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. IEEE Trans. Autom. Control, 35(3) : 289-298.

S. Meyn. (2008). Control techniques for complex networks. Cambridge University Press, New York.

S. P. Meyn. (1995). Transience of multiclass queueing networks via fluid limit models. Ann. Appl. Probab., 5(4) : 946-957.

A. Pukhalskij and A. Rybko. (2000). Nonergodicity of a queueing network under nonstability of its fluid model. Probl. Inf. Transm., 36(1) : 23-41.

A. Rybko and A. Stolyar. (1992). Ergodicity of stochastic processes describing the operation of open queueing networks. Probl. Inf. Transm., 28(3) : 199-220.

A. N. Rybko and A. L. Stolyar. (1993). On the ergodicity of random processes that describe the functioning of open queueing networks. Problems of Information Transmission 28 : 199-220.

T. I. Seidman. (1994). First come, first served can be unstable. IEEE Trans. Autom. Control, 39(10) : 2166-2171.

A. Stolyar. (1995). On the stability of multiclass queueing networks : A relaxed sufficient condition via limiting fluid processes. Markov Process. Relat. Fields, 1(4) : 491-512.

H. Q. Ye and H. Chen. (2001). Lyapunov method for the stability of fluid networks. Operations Research Letters, 28(3) : 125-136.

Chapitre 2

Stability condition of a priority queueing system

Ce chapitre a fait l'objet d'une publication dans le journal Mathematical Sciences And Applications E-Notes., Volume, 1(2),165-172, 2013.

Stability condition of a priority queueing system

Amina Angelika BOUCHENTOUF, Hanane SAKHI,
Department of Mathematics, Djillali Liabes university of Sidi Bel Abbes.
B. P. 89, Sidi Bel Abbes 22000, Algeria.
E-mail :bouchentouf_amina@yahoo.fr
Department of Mathematics, Sciences and Technology University of Oran : Mohamed Boudiaf, USTOMB.
B. P. 1505 EL-M'NAOUAR- Oran, Algeria.
E-mail :sakhi.hanane@yahoo.fr

Abstract
The present paper is devoted to the study of a priority multiclass system stability, where we establish a sufficient condition for the stability of a queueing network composed of N-units, $N \geq 3$ and N^2 classes (N classes at each unit) ; that is the stability under all work conserving disciplines with the priority discipline.
subclass [2000] : Primary 60K25 ; Secondary 68M20 ; Thirdly 90B22.

Keywords : Stability, fluid models, multiclass queueing networks, global stability, piecewise linear Lyapunov functions, linear Lyapunov functions.

2.1 Introduction

Queueing systems constitute a central tool in modelling and performance analysis of computer systems, communication systems, manufacturing systems and logistic systems.
An important tool for studying the stability of a queueing network is its corresponding fluid network, which is a continuous analog of the queueing network. An elegant theorem proposed by Rybko and Stolyar [14] then extended by Dai [4] states that a queueing network is stable if its corresponding fluid limit model or fluid network model is stable. This motivates the study of the stability of fluid networks.
Dai and Vande Vate [9, 7] characterized the global stability region of two-station fluid networks, in [5]the authors extended the methods used in [1, 2, 9, 7] to networks with three stations.
In this paper, we establish the global condition stability of a priority multiclass queueing networks composed of N units $N \geq 3$. using the peacewise linear Lyapunov function approch.
Determining the global stability region of such system is very difficult under bad disciplines, even if the traffic intensity at each unit of the network is less that one.
In fact we show that, with the usual traffic conditions

$$\rho = \left(\alpha \sum_{k:\sigma(k)=i} m_k < 1 \right), \text{ for } i = 1, 2, ..., N,$$

the following additional condition :

$$\alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} \leq 1$$

is sufficient to ensure global stability.

2.2 Model Description and Notations

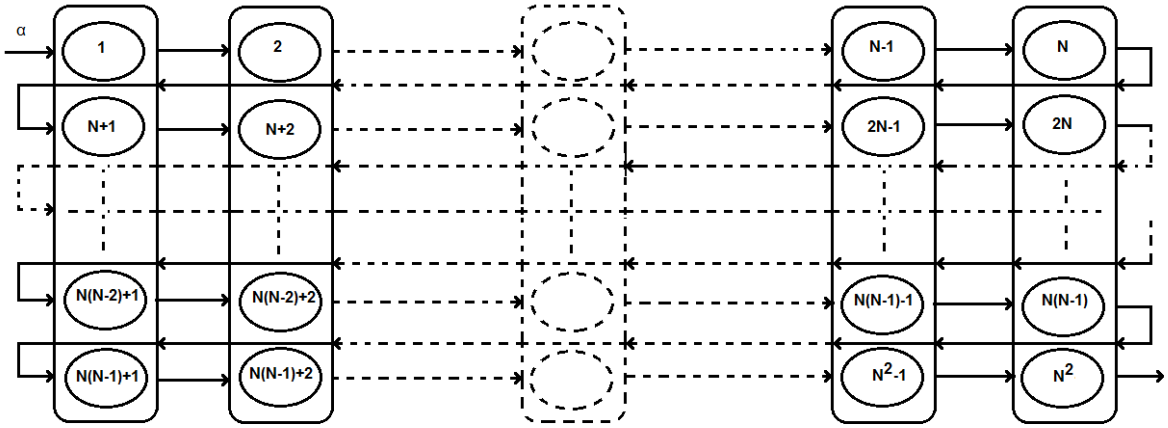

FIGURE 2.1 – Queueing system with N units and N^2 classes

We consider a model of a priority queueing system composed of N units with N^2 classes, at each unit a server may serve N classes of customers. Fluid comes from outside to the network at rate α per unit of time, requiring a service of mean m_1, then it aligns the second queue asking for a service of mean m_2, after that it removes to the third unit where it will be served another time, then to the fourth one, the fifth one..., until the Nth unit, where it will be served with mean m_N, afterward it returns to unit 1 and is again served by each unit. After being processed by the Nth units N^2 times it leaves definitively the system. The specific priority discipline under consideration is as follows : If the order of the station i, $i = \overline{1,N}$ is odd, the class k which belongs to this station has a priority over the above class which is the class $k-N$, and if the order of the station $i, i = \overline{1,N}$ is even, the class k which belongs to this station, has a priority over the bellow class which is the class $k+N$. Each fluid residing in buffer k, $k = 1, 2, ..., N^2$ is called class k fluid.

Let $Q_k(t)$ denote the fluid level in buffer k at time t, and $T_k(t)$, the cumulative time vector $\sigma(k)$ devotes the class k in the interval $[0,t]$.

We denote by $U_i(t) = t - \sum_{k:\sigma(k)=i} T_k(t)$ the cumulative idle time at unit i, $i = 1, 2, ...N$ in the interval $[0,t]$. The buffer levels $Q(.) = (Q_k(.))_{1\leq k\leq N^2}$ and the allocations $T(.) = (T_k(.))_{1\leq k\leq N^2}$ must satisfy :

$$Q_k(t) = Q_k(0) + \mu_{k-1}T_{k-1}(t) - \mu_k T_k(t), \quad t \geq 0, \quad \text{for} \quad k = 1, ..., N^2, \tag{2.1}$$

$$Q_k(t) \geq 0, \quad t \geq 0, \quad \text{for} \quad k = 1, ..., N^2, \tag{2.2}$$

$$T_k(.) \text{ is nondecreasing}, \quad k = 1, ..., N^2, \tag{2.3}$$

$$U_i(.) \text{ is nondecreasing}, \quad i = 1, ..., N, \tag{2.4}$$

where, $\mu_k = 1/m_k$ is the service rate for class k, $k = 1, 2, ..., N^2$, $\mu_0 = \alpha$ is the exogenous arrival rate and $T_0(t) = t$ models the exogenous arrival process, $\mu_k T_k(t)$ is the amount of fluid to have departed buffer k by time t.

Any solution $(Q(.), T(.))$ to (2.1)-(2.4) is a fluid solution. A fluid solution $(Q(.), T(.))$ satisfying :

$$\int_0^\infty Z_i(t) dU_i(t) = 0, \;\; i = 1, ..., N, \tag{2.5}$$

where

$$Z_i(t) = \sum_{k:\sigma(k)=i} Q_k(t), \;\; i = 1, ..., N, \tag{2.6}$$

is said to be non-idling or work conserving. Equations (2.1)-(2.5) define the fluid network under non-idling disciplines.

For any fluid solution $(Q(.), T(.))$, $Q(.)$ is differentiable for almost all t in $(0, \infty)$; t is regular point for fluid solution $(Q(.), T(.))$ if $T(.)$ is differentiable at t. For a differentiable function (at t) $h : [0, \infty) \to \mathbb{R}$, $\dot{h}(t)$ is the derivative of h at t. Notice that (2.5) is equivalent to the condition "$Z_i(t) > 0$ implies $\dot{U}_i(t) = 0$," for each regular point t.

It was shown in [2] that each fluid limit is a fluid solution satisfying (2.1)-(2.5). One of the particularly of non-idling discipline are the static buffer priority disciplines, which means that the server can only work on lower priority classes at a unit when the requirements of higher priority classes are satisfied. Each unit in our network serves N classes. Let π_i denote the high priority class at unit i under the static buffer priority discipline π. So, the specific priority discipline under consideration is as follows : If the order of the station i, $i = \overline{1, N}$ is odd, the class k which belongs to this station has a priority over the above class which the class $k - N$, and if the order of the station $i, i = \overline{1, N}$ is even, the class k which belongs to this station, has a priority over the bellow class which the class $k + N$. For instance $\pi_{\{13,9,5,1,2,6,10,14,15,11,7,3,4,8,12,16\}}$ denotes the static buffer priority discipline that gives in station 1 higher priority to class 13 over class 9, higher priority to class 9 over class 5 and higher priority to 5 over 1, in station 2 it gives higher priority to class 2 over 6, higher priority to 6 over 10 and higher priority to 10 over 14, and in station 3 it gives higher priority to 15 over 11, higher priority to 11 over 7 and higher priority to 7 over 3, and finally in station 4 it gives higher priority to 4 over 8, higher priority to 8 over 12 and higher priority 12 over 16.

With this notation, our fluid network under the static buffer priority π requires the additional equations :

$$\dot{T}_{\pi(i)}(t) = 1 \text{ if } Q_{\pi(i)}(t) > 0, \;\; i = 1, ..., N, \tag{2.7}$$

for each regular point t of $T(.)$. Any solution $(Q(.), T(.))$ to (2.1)-(2.5) and (2.7) is a fluid solution under the discipline π.

Definition 2.2.1 • *The fluid network is globally stable if there exists a time $\sigma > 0$ such that for each non-idling fluid solution $(Q(.), T(.))$ satisfying (2.1)-(2.5) with $|Q(0)| = 1$, $Q(t) = 0$ for all $t \geq \sigma$.*

• *The fluid network under a static buffer priority discipline π is stable if there exists a time $\sigma > 0$ such that for each fluid solution $(Q(.), T(.))$ satisfying (2.1)-(2.5) and (2.7) with $|Q(0)| = 1$, $Q(t) = 0$ for all $t \geq \sigma$.*

• *A fluid solution $(Q(.), T(.))$ is unstable if there is no $\sigma > 0$ such that $Q(t) = 0$ for all $t \geq \sigma$.*

• *For a given $\alpha > 0$, the global stability region of the fluid network is the set of positive service times $m = (m_k)$ for which the fluid network is globally stable.*

• *For a given $\alpha > 0$ and a static buffer priority discipline π, the stability region of the fluid network under the discipline is the set of positive services times $m = (m_k)$ for which the fluid network under the discipline is stable.*

We say that the usual condition is satisfied if the traffic intensity for each unit is less than 1 i.e.,

$$\rho = \alpha \sum_{k:\sigma(k)=i} m_k < 1, \text{ for each } \quad i = 1, ..., N. \tag{2.8}$$

At that moment, we say that the usual traffic conditions are satisfied.

Now, we define the following system of linear constraints, which is related to a piecewise linear Lyapunov function for our N-unit fluid network :

$$\alpha \sum_{k:\sigma(k)=i} x_k < x_k \mu_k, \quad i = 1, ..., N, \tag{2.9}$$

$$\left(\sum_{k:\sigma(k)=1} x_k \right) - x_1 \leq \sum_{k:\sigma(k)=N} x_k \tag{2.10}$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i \leq \sum_{k:\sigma(k)=i-1} x_k - x_i \quad i = 2, ..., N, \tag{2.11}$$

$$\sum_{k:\sigma(k)=i} x_k \leq \sum_{k:\sigma(k)=i-1} x_k \quad i = 2, ..., N. \tag{2.12}$$

The system of linear constraints (2.9)-(2.12) derived from our piecewise linear Lyapunov function provides conditions sufficient to ensure the global stability of the system.

2.3 Main result

In this section we characterize the global stability region using the piecewise linear Lyapunov functions. Given $x = (x_k) > 0$ and a fluid solution $Q(.)$, let

$$h_i(x, Q(t)) = \sum_{k:\sigma(k)=i} x_k Q_k^+(t), \quad i = 1, ..., N,$$

where $Q_k^+(t) = \sum_{l=1}^{k} Q_l(t)$, and $h(x, Q(t)) = \max\{h_i(x, Q(t))\}, \quad i = 1, ..., N$. So, $h(Q(t))$ is a convex, piecewise linear function of $Q(t) = (Q_k(t))$.

The piecewise linear function h is said to be Lyapunov function for the global stability of the fluid model if there exist $\varepsilon > 0$ such that for each non-idling fluid solution $(Q(.), T(.))$ satisfying (2.1)-(2.5),

$$\frac{dh(Q(t))}{dt} \leq -\varepsilon \tag{2.13}$$

for each time $t > 0$ that is regular for $T(.)$ and $h(Q(t))$ with $|Q(t)| > 0$.
Let $m > 0$ be a service time vector for which there is a piecewise Lyapunov function h satisfying (2.13). It follows from Lemma 2.2 of Dai and Weiss [9] that $h(Q(t)) = 0$ for all $t \geq h(Q(0))/\varepsilon$, or $Q(t) = 0$ for all $t \geq h(Q(0))/\varepsilon$.

The next lemma suggests a way in which to construct piecewise linear Lyapunov functions, it was introduced by Botvich and Zamyatin [2] for a two-station network. It was independently generalized by Dai and Weiss [8], Down and Meyn [9] and [5]. Using that result we get

Lemma 2.3.1 *Suppose there exists $x = (x_k) > 0$, $t_0 \geq 0$ and $\varepsilon > 0$ such that for each non-idling fluid solution $(Q(.), T(.))$ and each regular point $t > t_0$ of $T(.)$, the following hold for each $i = 1, 2, ..., N$*

$$\frac{dh_i(x, Q(t))}{dt} \leq -\varepsilon \ \textit{whenever} \ Z_i(t) > 0, \tag{2.14}$$

$$h_i(x, Q(t)) \leq \max\{h_j(x, Q(t)) : j \in \{1, ..., N\}, j \neq i\} \ \textit{whenever} \ Z_i(t) = 0, \tag{2.15}$$

$$\max\{h_j(Q(t)) : j \in \{1, ..., N\}, j \neq i\} \leq h_i(Q(t)) \ \textit{whenever} \ \sum_{j \neq i} Z_j(t) = 0. \tag{2.16}$$

Then h is a piecewise linear Lyapunov function.

Proposition 2.3.1 *If there exists $x = (x_k) > 0$ satisfying the linear constraints (2.9)-(2.12), then there exists $\varepsilon > 0$ such that (2.14)-(2.16) hold and hence, h is a piecewise linear Lyapunov function.*

Proof. Let $t_0 = 0$ and let $x = (x_k) > 0$ satisfying the linear constraints (2.9)-(2.12), define $\varepsilon > 0$ to be the minimum of the following terms $x_k\mu_k - \alpha\left(\sum_{k:\sigma(k)=i} x_k\right)$, $i = 1, ..., N$. Remark that the amount of fluids in buffers 1 through k is

$$Q_k^+(t) = Q_k^+(0) + \alpha t - \mu_k T_k(t).$$

Thus

$$h_1(x, Q(t)) = h_1(0) + \alpha t \left(\sum_{k:\sigma(k)=1} x_k\right) - \sum_{k:\sigma(k)=1} x_k\mu_k T_k(t)$$

and

$$\frac{dh_1(Q(t))}{dt} = \alpha(\sum_{k:\sigma(k)=1} x_k) - \sum_{k:\sigma(k)=1} x_k\mu_k \dot{T}_k(t)$$

If $Z_1(t) > 0$, it follows from (2.15), $\sum_{k:\sigma(k)=1} \dot{T}_k(t) = 1$, thus $\dot{h}_1(t) < -\varepsilon$.

We follow similar analysis for $i = \overline{2, N}$.

Next, we establish (2.15).

- When $Z_1(t) = 0$, equation (2.10) ensures that $h_1(Q(t)) < h_N(Q(t))$.
- When $Z_i(t) = 0,\ \ i = \overline{2, N}$, equation (2.11) ensures that $h_i(Q(t)) < h_{i-1}(Q(t))$.

Finally, we establish (2.16).

When $\sum_{j \neq i} Z_j(t) = 0,\ \ j = \overline{1, N}$, equations (2.11)-(2.12) ensures that $h_j(Q(t)) < h_i(Q(t))$.

The following result establishes a sufficient conditions to ensure global stability for our network.

Proposition 2.3.2 *If*

$$\alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} \leq 1, \tag{2.17}$$

$$\alpha \sum_{k:\sigma(k)=i} m_k < 1, \ \ i = 2, ..., N, \tag{2.18}$$

the fluid network is globally stable.

Proof. Let $(Q(.), T(.))$ be a non-idling fluid solution with $|Q(0)| = 1$. Let

$$f_1(t) = \sum_{k:\sigma(k)=1} \sum_{j=1}^{k} m_k Q_j(t)$$

be the total workload at the first unit at time t.

From (2.1) we get

$$f_1(t) = f_1(0) + \alpha t \left(\sum_{k:\sigma(k)=1} m_k \right) - \sum_{k:\sigma(k)=1} T_k.$$

For each regular t with $Z_1(t) > 0$, by (2.5), $\dot{f}_1(t) = \left(\alpha \sum_{k:\sigma(k)=1} m_k \right) - 1$. Since $\alpha \sum_{k:\sigma(k)=1} m_k < 1$, there is positive t_0 with

$$t_0 \leq \frac{f_1(0)}{1 - \alpha \sum_{k:\sigma(k)=1} m_k} \leq \frac{\sum_{k:\sigma(k)=1} k m_k}{1 - \alpha \sum_{k:\sigma(k)=1} m_k}$$

such that $Z_1(t_0) = 0$. Assume that (2.17) holds. We next show $Z_1(t) = 0$ for $t \geq t_0$. To see this, let

$$f_2(t) = \sum_{k:\sigma(k)=1} m_k Q_k(t)$$

be the immediate workload at unit 1. Then, from (2.1)-(2.4),

$$f_2(t) = f_2(0) + \alpha m_1 t - \sum_{k:\sigma(k)=1} T_k(t) + \sum_{k:\sigma(k)=1} m_k \mu_{k-1} T_{k-1}(t)$$

and, for any regular t with $f_2(t) > 0$,

$$\dot{f}_2(t) = \alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} - 1 \leq 0.$$

Thus f_2 is non-increasing. Since $f_2(t_0) = 0$, we have $Z_1(t) = 0$ for $t \geq t_0$.

Now, it remains to show that $\sum_{i=2}^{N} Z_i(t) = 0$ for each time $t \geq t_1 \geq t_0$ and that the network is globally stable. So, we consider times $t \geq t_0$ and specialize the lemma 2.3.1 to the case where $Z_1(t) = 0$ and $\dot{Q}_1(t) = ... = \dot{Q}_{N(N-1)+1}(t) = 0$. First, observe that since $Z_1(t) = 0$ for $t \geq t_0$, (2.14), (2.16) are satisfied for $i = 1$, Then, recalling that (2.16) implies (2.15) in our network, we see that we are left with the conditions :

$$\frac{dh_i(x, Q(t))}{dt} \leq -\varepsilon \text{ whenever } Z_i(t) > 0, \ \ i = 2, ..., N \tag{2.19}$$

$$\begin{aligned} &h_k(Q(t)) = max\{h_i(Q(t)), h_k(Q(t)), i \in \{1, ..., N\}\backslash\{k, j\}, \ \ k = 2, ..., N\} \\ &\leq h_j(Q(t)), \ \ j \in \{1, ..., N\}\backslash\{k, i\} \text{ whenever } \sum_{l=1}^{N\backslash\{j\}} Z_l = 0. \end{aligned} \tag{2.20}$$

After that, by Lemma 2.3.1, we can easily show that (2.19)-(2.20) and hence (2.14)-(2.16) hold if there exists $(x_2, x_3, ..., x_{N^2-1}, x_{N^2}) > 0$ satisfying

$$\alpha \sum_{k:\sigma(k)=i} x_k < \mu_k x_k, \ \ i = 2, ..., N \tag{2.21}$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=N} x_k, \ \ i = 2, ..., N-1 \tag{2.22}$$

$$\sum_{k:\sigma(k)=i} x_k < \sum_{k:\sigma(k)=i-1} x_k, \ \ i = 2, ..., N \tag{2.23}$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=i-1} x_k, \ \ i = 3, ..., N \tag{2.24}$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=i-1} x_k - x_i, \ \ i = 3, ..., N \tag{2.25}$$

Then, given $(x_1, x_2, ..., x_{N^2}) > 0$, let

$$y_i^k = \frac{x_k}{\sum_{k:\sigma(k)=i} x_k}, i = 1, ..., N$$

And $(x_1, ..., x_{N^2}) > 0$ satisfies (2.9)-(2.12) iff $(y_1, ..., y_N, x_{N+1}, ..., x_{N^2}) > 0$ satisfies

$$\alpha m_k < y_i^k, \;\; k = 1, ..., N^2, \;\; i = 1, ..., N. \tag{2.26}$$

So, there exists $x > 0$ satisfying (2.21)-(2.25) if and only if the usual traffic condition (2.18) at units $2, 3, ..., N$ hold. Therefore, the proposition follows from Lemma 2.3.1. □

Bibliographie

[1] D. Bertsimas, D.Gamarnik and J.N.Tsitsiklis., Stability conditions for multiclass fluid queueing networks. IEEE Trans. Automat. Control. **41** (1996), no. 11, 1618-1631.

[2] D.D. Botvich and A.A. Zamyatin., Ergodicity of conservative communication networks. Rapport de recherche, 1772, INRIA, (1992).

[3] H. Chen. Fluid approximations and stability of multiclass queueing networks I : Work-conserving disciplines., Annals of Applied Probability. **5** (1995), 637-665.

[4] J. G. Dai., On possitive Harris recurrence of multiclass queueing networks : a unified approach via fluid limit models. Mathematics and its applications. **71** (1995), 71-90.

[5] J. G. Dai, J. J. Hasenbein and J. Vande Vate., stability of a three-station fluid network. Queueing Systems. **33** (1999), 293-325.

[6] J. G. Dai and J. Vande Vate., Global stability of two-station queueing networks. Lecture Notes in Statistics. Columbia University, New York, Springer-Verlag, **117** (1996), 1-26.

[7] J. G. Dai and J. Vande Vate., The stability of two-station fluid networks. Operations Research. **48** (2000), 721-744.

[8] J. G. Dai and G. Weiss., Stability and instability of fluid models for re-entrant lines. Mathematics of operations Research. **21** (1996), 115-134.

[9] D. Down and S.P. Meyn., Piecewise linear test functions for stability and instability of queueing networks. Queueing Systems. **27** (1997), 205-226.

[10] A. N. Rybko and A. L. Stolyar., On the ergodicity of random processes that describe the functioning of open queueing networks. Problems on Information Transmission. **28** (1992), no. 3, 3-26.

Chapitre 3

Stabilizing priority fluid queueing network model

Ce chapitre a fait l'objet d'une publication dans le journal Acta Univ. Sapientiae, Mathematica., Volume, 6(2), 146-161, 2014.

Stabilizing priority fluid queueing network model

Amina Angelika BOUCHENTOUF, Hanane SAKHI,
Department of Mathematics, Djillali Liabes university of Sidi Bel Abbes.
B. P. 89, Sidi Bel Abbes 22000, Algeria.
E-mail :bouchentouf_amina@yahoo.fr
Department of Mathematics, Sciences and Technology University of Oran : Mohamed Boudiaf, USTOMB.
B. P. 1505 EL-M'NAOUAR- Oran, Algeria.
E-mail :sakhi.hanane@yahoo.fr

Abstract
The aim of this paper is to establish the stability of fluid queueing network models under priority service discipline. To this end, we introduce a priority fluid multiclass queueing network model, composed of N stations, $N \geq 3$ and N^2 classes (2 classes at each station) ; where in the system, each station may serve more than one job class with differentiated service priority, and each job may require service sequentially by more than one service station. In this paper the fluid model approach is employed in the study of the stability.

subclass [2000] :60K25 ; 68M20 ; 90B22.

Keywords : Stability, fluid models, multiclass queueing networks, fluid approximation.

3.1 Introduction

Stochastic processing networks arise as models in manufacturing, telecommunications, computer systems and service industry. Common characteristics of these networks are that they have entities, such as jobs, customers or packets, that move along routes, wait in buffers, receive processing from various resources, and are subject to the effects of stochastic variability through such quantities as arrival times, processing times, and routing protocols. Networks arising in modern applications are often highly complex and heterogeneous. Typically, their analysis and control present challenging mathematical problems. One approach to these challenges is to consider approximate models.
In the last 15 years, significant progress has been made on using approximate models to understand the stability and performance of a class of stochastic processing networks called open multiclass HL queueing networks. HL stands for a non-idling service discipline that is head-of-the-line, i.e., jobs are drawn from a buffer in the order in which they arrived. Examples of such disciplines are FIFO and static priorities. First order (functional laws of large numbers) approximations called fluid models have been used to study the stability of these networks, and second order (functional central limit theorem) approximations which are diffusion models, have been used to analyze the performance of heavily congested networks.

The development of the fluid approach was inspired by the studies of some counter-examples in Kumar and Seidman [11], Rybko and Stolyar [14] and Bramson [1], etc., where the multiclass queueing networks are not stable even when the traffic intensity of each station in the network is less than one. An elegant result of the fluid model approach was proposed first in Rybko and Stolyar [14] and then generalized and refined by Dai [6], Chen [2], Dai and Meyn [8], Stolyar [15] and Bramson [1]. It states

that a queueing network is stable if its corresponding fluid network model is stable. Partial converse to this result is also given in Meyn [12], Dai [7] and Puhalskii and Rybko [13]. Heng Quing Ye [10] used Kumar-Rybko-Seidman-Stolyar network for establishing the stability of fluid queueing network.

In this paper, we concentrate with the capacity of some large classes of fluid multiclass queueing networks under priority service discipline. Specifically, we establish a stability condition of some heterogenous priority fluid networks with N stations and $2N$ job classes, where in the system, each station may serve more than one job class with differentiated service priority, and each job may require service sequentially by more than one service station. So, in our case, the network performance is improved even when more workloads are admitted for service. To stabilize our networks a number of stations should be added, these later act as regulators for the systems, adding these stations is not random, it depends essentially on higher and lower priority job classes (many-to-one mapping) and on the number of stations in the network. The fluid model approach is employed to proof the stability.
The outline of the paper is as follows : At first (section 2) we describe priority fluid multiclass queueing models, and present a powerful result on the stability of such systems given by Chen and Zhang [5], after that (section 3) we introduce a modified networks and present their stability conditions (theorems 3.3.1 and 3.3.2), and finally we conclude this paper with a short conclusion.

3.2 N-stations priority fluid multiclass network models

We describe N-stations priority fluid queueing network models as

$(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$. Specifically, the fluid network consists of J stations (buffers) $(J = N)$ indexed by $j \in \mathcal{J} = \overline{1, N}$, serving K, $K = 2N$ fluid (customer) classes indexed by $k \in \mathcal{K} = \overline{1, 2N}$. A fluid class is served exclusively at one station, but one station may serve more than one fluid classes. $\sigma(.)$ denotes a many-to-one mapping from $\mathcal{K}$ to $\mathcal{J}$, with $\sigma(k)$ indicating the station at which a class k fluid is served. A class k fluid may flow exogenously into the network at rates λ_1 and λ_{N+1}, (≥ 0), then it is served at station $\sigma(k)$, with mean service time $m_k = 1/\mu_k$, $k = \overline{1, 2N}$ and after being served, a fraction p_{kl} of fluid turns into a class l fluid, $l \in \mathcal{K}$, and the remaining fraction, $1 - \sum_{l=1}^{K} p_{kl}$ flows out of the network. Let $C(j)$ be the set of classes that reside in station j, alternatively, we denote by a $J \times K$ matrix $C = (c_{ij})_{J \times K}$, known as the constituent matrix where $c_{jk} = 1$ if $\sigma(k) = j$, and $c_{jk} = 0$ otherwise.
Let $Q_k(t)$ indicates the number of class k customers in the network at time t, $(Q(0) = Q_k(0))$ and $\lambda = (\lambda_k)$ two K-dimensional nonnegative vectors. $P = (p_{kl})_{K \times K}$ a stochastic matrix with spectral radius strictly less that one, $\mu = (\mu_k)$ a K-dimensional positive vector.
The vectors $Q(0)$ are referred to as initial fluid level vector, λ to the exogenous inflow rate vector, μ to the processing rate vector, Matrix P is referred to as flow transfer matrix.

When station $\sigma(k)$ devotes its full capacity to serving class k fluid (assuming that it is available to be served), it generates an outflow of class k fluid at rate $\mu_k > 0$, $k \in \mathcal{K}$. Among classes, fluid follows a priority service discipline, which is again described by a one-to-one mapping π from $\{1, ..., K\}$ onto itself. Specifically, a class k has priority over a class l if $\pi(k) < \pi(l)$ and $\sigma(l) = \sigma(k)$, then class k job can not

be served at station $\sigma(k)$ unless there is no class l job.

So, our multiclass fluid network consists of N stations and $2N$ job classes. Assume that the arrival process of class k, $k = \overline{1, 2N}$, customers arrive to the system following a Poisson process with arrival rates $\lambda_1 \geq 0$ and $\lambda_{N+1} \geq 0$, the service time for each class k customer is exponentially distributed with mean service time $m_k > 0$. We also assume that all the inter-arrival times and service times are independent.

To describe the dynamics of the fluid network, we introduce the K-dimensional fluid level process $\overline{Q} = \{\overline{Q}(t), \ t \geq 0\}$, whose kth component $\overline{Q}_k(t)$ denotes the fluid level of class k at time t; the K-dimensional time allocation process $\overline{T} = \{\overline{T}(t), t \geq 0\}$, whose kth component $\overline{T}_k(t)$ denotes the total amount of time that station $\sigma(k)$ has devoted to serving class k fluid during the time interval $[0, t]$, and the K-dimensional unused capacity process $\overline{Y} = \{\overline{Y}(t), t \geq 0\}$, whose kth component $\overline{Y}_k(t)$ denotes the (cumulative) unused capacity of station $\sigma(k)$ during the time interval $[0, t]$ after serving all classes at station $\sigma(k)$ which have a priority no less than class k. We denote by D the K-dimensional diagonal matrix whose kth element is μ_k, and e is a K-dimensional vector with all components being one. Let

$$H_k = \{l : \sigma(l) = \sigma(k), \pi(l) \leq \pi(k)\}$$

be the set of indices for all classes that are served at the same station as class k and have a priority no less than that of class k. Note that $k \in H_k$ by definition. Then the dynamics of the fluid network model can be described as follows.

$$\overline{Q}(t) = \overline{Q}(0) + \lambda t - (I - P')D\overline{T}(t) \geq 0, \tag{3.1}$$

$$\overline{T}(\cdot) \quad \text{is nondecreasing with } \overline{T}(0) = 0, \tag{3.2}$$

$$\overline{Y}_k(t) = t - \sum_{l \in H_k} \overline{T}_l(t) \ \text{ is nondecreasing, } \ k \in \mathcal{K}, \tag{3.3}$$

$$\int_0^\infty \overline{Q}_k(t) d\overline{Y}_k(t) = 0, \ \ k \in \mathcal{K}. \tag{3.4}$$

Let

$$Q_k(t) = Q_k(0) + \lambda_k t + \sum_{l=1}^{K} p_{lk}\mu_l T_l(t) - \mu_k T_k(t) \geq 0, k = 1, ..., K, \tag{3.5}$$

be the kth coordinate of the flow balance relation (3.1).

The equation (3.1) is the equivalent relation between the time allocation process $T(\cdot)$ and the unused capacity process $Y(.)$. The relation (3.4) means that at any time t, there could be some positive remaining capacity (rate) for serving those classes at station $\sigma(k)$ having a strictly lower priority than class k, only when the fluid levels of all classes in H_k (having a priority no less than k) are zero.

A pair $(\overline{Q}, \overline{T})$ (or equivalently $(\overline{Q}, \overline{Y})$) is said to be a fluid solution if they jointly satisfy (3.1)-(3.4). For convenience, we also call $\overline{Q}$ a fluid solution if there is a $\overline{T}$ such that the pair $(\overline{Q}, \overline{T})$ is a fluid solution. The

fluid network $(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$ is said to be stable if there is a time $\tau \geq 0$ such that $\overline{Q}(\tau + \cdot) \equiv 0$ for any fluid solution $\overline{Q}$ with $\|\overline{Q}(0)\| = 1$; and it is said to weakly stable if $\overline{Q}(\cdot) = 0$ for any fluid solution $\overline{Q}$ with $\overline{Q}(0) = 0$. The processes $\overline{Q}$, $\overline{Y}$, and $\overline{T}$ are Lipschitz continuous, and hence are differentiable almost everywhere on $[0, \infty)$, this well-known property will be used in this paper.

It is well known that the queue length process $Q(t)$ is a continuous time Markov chain under the Poisson arrival and exponential service assumptions. We say that the network $(\mathcal{J}, \mathcal{K}, \lambda, m, C, P, \pi)$ is stable if the Markov chain $Q(t)$ is positive recurrent. It is well know that the Markov chain $Q(t)$ is positive recurrent only if the traffic intensity for each station is less than one, i.e., $\rho_j < 1$ (ρ_j is the jth component of ρ; a traffic intensity for station j) for all $j \in \mathcal{J}$, or in short, $\rho < e$ where e is a J-dimensional vector with all components being ones.

The expected stationary total queue length $\overline{Q}$ is defined as

$$\overline{Q} = \lim_{t \to \infty} \mathbb{E}\left[\sum_{k \in \mathcal{K}} Q_k(t)\right].$$

The queue length $\overline{Q}(t)$ is a finite if and only if the queue length process Q is positive recurrent.

Chen and Zhang [5] gave a very important result on the stability of priority fluid queueing systems, authors established the sufficient condition for the stability based on the existence of a linear Lyapunov function, this later (sufficient condition) gave the the necessary and sufficient condition for the stability. Their result is presented in the theorem 3.2.1, in order to state it we need some additional assumptions :

Let

$$h(k) = \begin{cases} \arg\max\{\pi(l) : l \in H_k^+\} & \text{if } H_k^+ \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

with $H_k^+ = H_k \backslash \{k\}$, in words ; if k is not the highest priority class at station $\sigma(k)$, then $h(k)$ is the index for the class which has the next higher priority than class k at station $\sigma(k)$, otherwise $h(k) = 0$.

$$\theta = \lambda - (I - P')\mu_H^0, \tag{3.7}$$

where $\mu_H^0 = De_H^0$, $(e_H^0 = (e_1^0, ..., e_K^0)')$ is a K-dimensional vector with $e_k^0 = 1$ if $H_k^+ = \emptyset$ and $e_k^0 = 0$ otherwise.

$$R = (I - P')D(I - B), \tag{3.8}$$

where $B = (b_{lk})$ is the $K \times K$ matrix with $b_{lk} = 1$ if $k = h(l)$, and $b_{lk} = 0$ otherwise, $(l, k = 1, ...K)$.

And let

$$\rho = CD^{-1}(I - P')^{-1}\lambda. \tag{3.9}$$

be the traffic intensity of the queueing network.

Theorem 3.2.1 *[5] Consider a fluid network (λ, μ, P, C) under priority service discipline π. Let vector θ and matrix R be as defined in (3.7),(3.8) respectively. Assume that $\rho < e$. Then the fluid network is stable if there exist a K-dimensional vector $h \geq 0$ such that for any given partition a and b of $\mathcal{K}$ satisfying if class $l \in a$, then each class k with*

$$\sigma(k) = \sigma(l) \ \text{ and } \ \pi(k) > \pi(l) \ \text{ is also in } a \tag{3.10}$$

we have

$$h'_a(\theta_a + R_{ab}x_b) < 0 \tag{3.11}$$

for $x_b \in \mathcal{S}_b := \{u \geq 0 : \theta_b + R_b u = 0 \ \text{ and } \ u \leq e\}$ when $b \neq \emptyset$, and $x_b = 0$ when $b = \emptyset$. The inequality (3.11) is omitted to hold by default when $S_b = \emptyset$.

Set a includes all classes which have a zero unused capacity rate and set b includes all classes which have a positive unused capacity rate at time t.

3.3 Main Result

In this paper, we present two theorems, we provide the proof of the first theorem, while the proof of the second one is omitted since it is similar to the former one.

3.3.1 Stabilizing N-stations priority fluid queueing network with some additional stations

Our multiclass queueing network consists of N stations and $2N$ job classes. Assume that the arrival process of class k; $k = \overline{1, 2N}$ customers arrive to the system following a Poisson process with arrival rates λ_1 and λ_{N+1} (≥ 0) to station 1 and $N+1$ respectively, the service time for each class k customer is exponentially distributed with mean service time $m_k > 0$: We also assume that all the inter-arrival times and service times are independent.

Suppose that each even class at station $i = \overline{1, N}$ has higher priority.ð

We modify our network such that if it is composed of an even number of stations we add N additional ones otherwise we add $(N-1)$, the explanation of this choice will be given in the rest of the paper, the modified network is illustrated in Figures 3.1 and 3.2 ; the additional stations are named station $N+1$,...,station $2N$, (N : even) (resp. station $N+1$,...,station $2N-1$ (N : odd)).

Theorem 3.3.1 *Suppose $\rho < e$, equation (3.11) not satisfied.*

If

$$\lambda_{k_1} > (1 - m_{k'_1}/m_{k''_1})/m_{k_1}. \tag{3.12}$$

Then the queue length process $Q(\cdot)$ is positive recurrent.

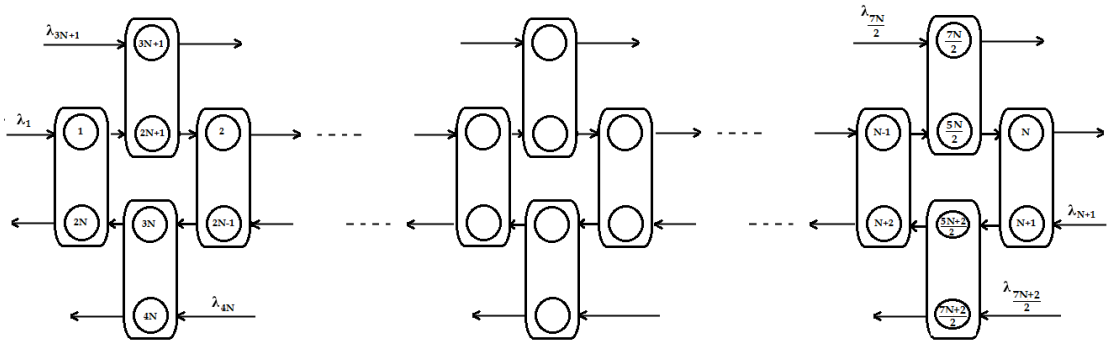

FIGURE 3.1 – 2N-stations priority fluid queueing Network

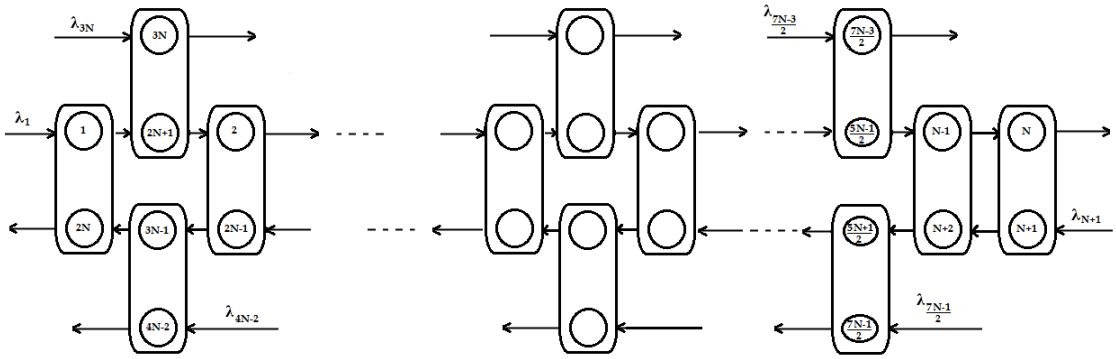

FIGURE 3.2 – 2N-1-stations priority fluid queueing Network

λ_{k_1} *(resp.* m_{k_1} *) is the exogenous arrival rate (resp. the mean service time) of higher priority fluid class of additional stations* $i = \overline{N+1, 2N}$, *(*N *: even) (resp.* $i = \overline{N+1, 2N-1}$, *(*N *: odd)), such that* $k_1 = \overline{3N+1, 4N}, (N : even)$ *(resp.* $k_1 = \overline{3N, 4N-2}$, *(*$N$ *: odd)).*

$m_{k_1'}$ *is the mean service time of lower priority fluid class of additional stations* $i = \overline{N+1, 2N}$, *(*N *: even) (resp.* $i = \overline{N+1, 2N-1}$, *(*N *: odd)).*

$m_{k_1''}$ *is the mean service time of higher priority fluid class of the original network.*

With

$$m_{k_1'} = \begin{cases} m_{k_1-N}, & k_1 = \overline{3N+1, 4N}, \quad (N : even), \\ m_{k_1-(N-1)}, & k_1 = \overline{3N, 4N-2}, \quad (N : odd). \end{cases}$$

$$m_{k_1''} = \begin{cases} m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N}{2}}, \ j_1 = \overline{1, \frac{N}{2}} \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+2}{2}, k_1-N}, \ j_1 = \overline{2, N} \end{cases} \quad (N : even),$$

$$m_{k_1''} = \begin{cases} m_{k_1'-(2N-j_1)}, & k_1' = \overline{2N+1, \frac{5N-1}{2}}, \ j_1 = \overline{1, \frac{N-1}{2}} \\ m_{(k_1-k_1')+j_1}, & k_1' = \overline{\frac{5N+1}{2}, k_1-(N-1)}, \ j_1 = \overline{4, N+1} \end{cases} \quad (N : odd).$$

Where for each k_1' *it corresponds* k_1 *and* j_1, *(*j_1 *is an even number).*

Via this theorem, we show that when the arrival rates of some job classes is reduced, the performance of the queue will worse. In Chen and Yao [3] and dai [7], it was shown that to prove the stability of a queueing model, it is sufficient to study the stability of its corresponding fluid queueing model, our prove is based on this result. To understand better the phenomenon, let us examine the dynamics of the original network with no initial job. When the higher priority job classes are being served, the lower priority ones are in standby, waiting for service, (class 1 jobs can not move to class 2 and 2 cannot move to 3,... for further services, and vice versa). So, these classes will never be served at the same time and in effect form a virtual stations (Dai and Vande Vate [9]). Therefore, the total nominal traffic intensity for these classes together, i.e., the virtual stations, should not exceed one for the network to be stable. The similar argument also yields that the network is unstable when the nominal traffic intensity for the virtual stations exceed one, i.e., the condition (3.11) is not satisfied. Now consider the modified network. The additional classes act as regulators that regulate the traffics in the network so as to stabilize the network. When the workloads of classes k_1 (k_1; defined in the theorem) are light, much service capacity of the additional stations are left to classes k_1' (k_1'; defined in the theorem) and hence these later do not hold back the traffics to avoid building up of job queues at priority classes of the original network. Thus, the virtual stations effect prevails and the network is still unstable when the condition (3.11) is not satisfied. However, when the workloads of classes k_1 are heavy enough such that the condition (3.12) holds, the service for lower priority classes at additional stations is in effect slowed down and the traffics in the original network are held back (these classes will not mutually block their services). Finally, the virtual station effect is avoided and the modified network is thus stabilized.

The dynamics of the our modified fluid network model can be described as follows.

$$\overline{Q}_{k_1}(t) = \overline{Q}_{k_1}(0) + \lambda_{k_1} t - \mu_{k_1} \overline{T}_{k_1}(t) \geq 0, \tag{3.13}$$

$k_1 = 1, N+1, \overline{3N+1, 4N}, (N : even), (resp. k_1 = 1, N+1, \overline{3N, 4N-2})(N : odd)$,

$$\overline{Q}_k(t) = \overline{Q}_k(0) + \mu_l \overline{T}_l(t) - \mu_k \overline{T}_k(t) \geq 0, \tag{3.14}$$

(k, l) = two successive job classes, such that the k^{th} class is the arriving l^{th} class

$$\overline{T}_k(\cdot) \quad \text{is nondecreasing with } \overline{T}_k(0) = 0, \tag{3.15}$$

$$k = \overline{1, 4N}, (N : even)(resp. k = \overline{1, 4N-2}, (N : odd))$$

$$\begin{cases} \overline{Y}_{k_1}(t) = t - \overline{T}_{k_1}(t), \\ \overline{Y}_{k_1''}(t) = t - \overline{T}_{k_1''}(t), \end{cases} \quad \text{is non decreasing,} \tag{3.16}$$

$$\overline{Y}_k(t) = t - \overline{T}_l(t) - \overline{T}_k(t) \text{ is nondecreasing,} \tag{3.17}$$

(k, l)=(lower priority job class, higher priority job class) at station i, $i = \overline{1, 2N}$, $(N : even)$ (resp. $i = \overline{1, 2N-2}$, $(N : odd)$),

$$\int_0^\infty \overline{Q}_k(t) d\overline{Y}_k(t) = 0, \quad k = \overline{1, 4N}(N : even) \ (resp. \ k = \overline{1, 4N-2}(N : odd)). \tag{3.18}$$

The stability study of the modified fluid network will be done in three steps.

1. First step. We prove that there exists a time $\tau_1 \geq 0$ such that

$$\overline{Q}_{k_1}(t) = 0, \quad \text{for any } t \geq \tau_1, \tag{3.19}$$

with $k_1 = \overline{3N+1, 4N}$, (N : even), (resp. $k_1 = \overline{3N, 4N-2}$, (N : odd)).

If $\dot{\overline{Q}}_{k_1}(t) > 0$, then we have by equation (3.18)

$$\dot{\overline{Y}}_{k_1}(t) = 0, \tag{3.20}$$

then by conditions (3.16) and (3.20)

$$\dot{\overline{T}}_{k_1}(t) = 1, \tag{3.21}$$

then by (3.13) and (3.21), we get

$$\dot{\overline{Q}}_{k_1}(t) = \lambda_{k_1} - \mu_{k_1}, \tag{3.22}$$

Note that the condition $\rho < e$ implies $\lambda_{k_1} < \mu_{k_1}$. Let $\tau_1^{(l)} = \dot{\overline{Q}}_{k_1}(0)/(\mu_{k_1} - \lambda_{k_1})$, $l = \overline{1, \frac{N}{2}}$, ($N$: even) (resp. $l = \overline{1, \frac{N-1}{2}}$, ($N$: odd)). Then, we have

$$\overline{Q}_{k_1}(t) = 0 \text{ for any } t \geq \tau_1^{(l)}. \tag{3.23}$$

Letting $\tau_1 = \max(1/\mu_{k_1} - \lambda_{k_1})$, we have that $\tau_1 \geq \max(\tau_1^{(l)})$, (each l corresponds to k_1) under the assumption $\|\overline{Q}(0)\| = 1$. Now, the conclusion (3.23) leads to the claim (3.19).

2. Second step. We prove that there exists a time $\tau_2 \geq \tau_1$ such that

$$\overline{Q}_{k_1''}(t) = 0, \text{for any } t \geq \tau_2, \tag{3.24}$$

where k_1'' is the higher priority job class at station i, $i = \overline{1, N}$.

$$k_1'' = \begin{cases} k_1' - (2N - j_1), & k_1' = \overline{2N+1, \frac{5N}{2}}, \ j_1 = \overline{1, \frac{N}{2}} \\ (k_1 - k_1') + j_1, & k_1' = \overline{\frac{5N+2}{2}, k_1 - N}, \ j_1 = \overline{2, N} \end{cases} \quad \text{(N : even)},$$

$$k_1'' = \begin{cases} k_1' - (2N - j_1), & k_1' = \overline{2N+1, \frac{5N-1}{2}}, \ j_1 = \overline{1, \frac{N-1}{2}} \\ (k_1 - k_1') + j_1, & k_1' = \overline{\frac{5N+1}{2}, k_1 - (N-1)}, \ j_1 = \overline{4, N+1} \end{cases} \quad \text{(N : odd)}.$$

Under the condition (3.19), we have $\dot{\overline{Q}}_{k_1}(t) = 0$, and then $\dot{\overline{T}}_{k_1}(t) = \lambda_{k_1} m_{k_1}$, $k_1 = \overline{3N+1, 4N}, (N : even)(resp. k_1 = \overline{3N, 4N-2}, (N : odd))$, for all time $t \geq \tau_1$. Combined with (3.17), this gives rise to

$$\dot{\overline{Y}}_{k_1'}(t) = t - \dot{\overline{T}}_{k_1'}(t) - \dot{\overline{T}}_{k_1}(t) \geq 0,$$

k_1' are classes of lower priority at additional stations,

$$k_1' = \begin{cases} k_1 - N, & k_1 = \overline{3N+1, 4N}, \quad (\text{N : even}), \\ k_1 - (N-1), & k_1 = \overline{3N, 4N-2}, \quad (\text{N : odd}). \end{cases}$$

and

$$\dot{\overline{T}}_{k_1'}(t) \leq 1 - \dot{\overline{T}}_{k_1}(t) = 1 - \lambda_{k_1} m_{k_1}, \text{ for any } t \geq \tau_1 \tag{3.25}$$

Then,

$\dot{\overline{Q}}_{k_1''}(t) = \mu_{k_1'} \dot{\overline{T}}_{k_1'}(t) - \mu_{k_1''} \dot{\overline{T}}_{k_1''}(t) \leq \mu_{k_1'}(1 - \lambda_{k_1} m_{k_1}) - \mu_{k_1''} < 0$, where for each k_1' it corresponds k_1'' for any $t \geq \tau_2$, where the last inequality is implied by the assumption that

$$\lambda_{k_1} > (1 - m_{k_1'}/m_{k_1''})/m_{k_1}.$$

Let $\tau_2^{(l)} = \frac{\dot{\overline{Q}}_{k_1''}(\tau_1)}{\mu_{k_1''} - \mu_{k_1'}(1-\lambda_{k_1} m_{k_1})}$, $l = \overline{1, N}$ $(N : even)$, (resp. $l = \overline{1, N-1}$ (N :odd)). Then, we have

$$\overline{Q}_{k_1''}(t) = 0 \text{ for any } t \geq \tau_2^{(l)}. \tag{3.26}$$

Let

$$\tau_2 = \max\left(\frac{1 + \Theta\tau_1}{\mu_{k_1''} - \mu_{k_1'}(1 - \lambda_{k_1} m_{k_1})}\right)$$

with Θ being the Lipschitz constant for the fluid level process $\overline{Q}(t)$. Then we have that $\tau_2 \geq \max(\tau_2^{(l)})$. Now, the conclusion (3.26) implies the claim (3.24).

Before to pass to the last step, we prove separately that $\overline{Q}_{N+1}(t) = 0$ for any $t \geq \tau_2$, " the case of network with even number of stations".
If $\dot{\overline{Q}}_{N+1}(t) = 0$, this implies $\dot{\overline{Y}}_{N+1}(t) = 0$, which in turn implies that $\dot{\overline{T}}_{N+1}(t) = 1$, then we have $\dot{\overline{Q}}_{N+1}(t) = \lambda_{N+1} - \mu_{N+1}$, with $\lambda_{N+1} < \mu_{N+1}$ (since $\rho < e$). So there exists $\tau_2' = \dot{\overline{Q}}_{N+1}(0)/\mu_{N+1} - \lambda_{N+1}$, such that $\dot{\overline{Q}}_{N+1}(t) = 0$ for any $t \geq \tau_2$.

Third step. We prove that there exists a time $\tau \geq \tau_2 (\geq 0)$ such that

$$\overline{Q}_l(t) = 0, \text{ for } t \geq \tau, \tag{3.27}$$

l represents job classes of lower priority at station $i = \overline{1, 2N}$ (N :even) (resp. $i = \overline{1, 2N-1}$ (N :odd)

which together with equations (3.19) and (3.24) implies

$$\overline{Q}(t) = 0 \text{ for } t \geq \tau.$$

Let

$$\overline{W}_i(t) = (\lambda_1 m_{l_1} + \lambda_{N+1} m_{l_2})t - \sum_{k:\sigma(k)=i} \overline{T}_k(t), \; i = \overline{1, N},$$

with $l_1 = \overline{1, N}$ and $l_2 = \overline{N = 1, 2N}$ job classes at the same station in the original network.

$\overline{W}_{i'}(t)=$

$$\begin{cases} \lambda_1 m_{k_1'} t - \overline{T}_{k_1'}(t), & k_1^{'} = \overline{2N+1, \frac{5N}{2}}, N : even, (resp. k_1^{'} = \overline{2N+1, 2N+\frac{5N-1}{2}}, N : odd) \\ \lambda_{N+1} m_{k_1'} t - \overline{T}_{k_1'}(t), & k_1^{'} = \overline{\frac{5N+2}{2}, 3N}, N : even, (resp. k_1^{'} = \overline{\frac{5N+1}{2}, 3N-1}, N : odd) \end{cases}$$

for $\tau \geq \tau_2$. Here $\overline{W}(t)$ can be explained as the immediately workload in the system at time t. Define

$$f_i(t) = k_1' \overline{W}_i(t), \text{ with } k_1' \text{ a lower priority job class in the additional stations.}$$

$$f_{i'}(t) = k_1'' \overline{W}_{i'}(t), \text{ with } k_1'' \text{ a higher priority job class in the original network.}$$

For each i (resp. i') it corresponds k_1' (resp. k_1'').

Then, it is direct to verify that, for $t \geq \tau_2$,

$$\dot{f}_i(t) < 0 \text{ if } \dot{\overline{Q}}_i(t) > 0, \text{ for } i = \overline{1, 4N}, (N : even), (resp. i = \overline{1, 4N-2}, (N : odd))$$

And

$$f_1(t) \leq f_N(t) \text{ if } \overline{Q}_1(t) = 0,$$

$$f_i(t) \leq f_{i-1}(t) \text{ if } \overline{Q}_i(t) = 0, \ i = \overline{2, 3N}, (N : even)(resp. i = \overline{2, 3N-1}, (N : odd))$$

$$f_j(t) \leq f_i(t) \text{ if } \sum_{j \neq i} \overline{Q}_j(t) = 0, \ j = \overline{1, 3N}, (N : even)(resp. j = \overline{1, 3N-1}, (N : odd))$$

$$f_N(t) \leq f_N(t) \text{ if } \overline{Q}_{3N}(t) = 0, (N : even)(resp. \overline{Q}_{3N-1}(t) = 0, (N : odd))$$

Now applying the piecewise linear Lyapunov function approach for the multiclass fluid network model described in Theorem 3.1 of Chen and Ye [4], we obtain the conclusion (3.27).

3.3.2 Stabilizing N-stations priority fluid queueing networks with N additional stations

Our N-stations multiclass queueing network is the same is above. Suppose in this case that the higher priority is devoted to classes $N, \overline{N+2, 2N}$.

We modify our network by adding N stations, (see Figure 3.3), compared to the original network, there are N additional stations, namely the station $N+1$,...,station $2N$, such that ; $3N+1$ job class has high priority at station $N+1$, and $\overline{3N+2, \ 4N}$ job classes have higher priority at stations $\overline{N+2, 2N}$. Now, let us introduce the second main result.

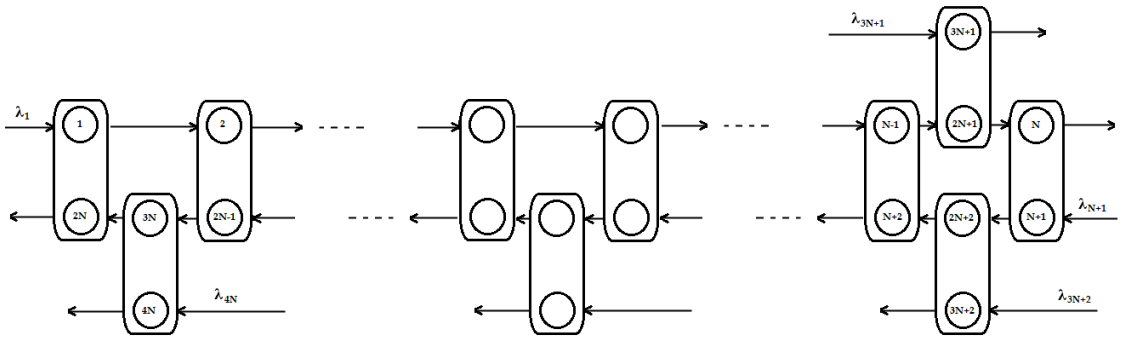

FIGURE 3.3 – 2N-fluid priority queueing Network

Theorem 3.3.2 *Suppose $\rho < e$ holds, equation (3.11) not satisfied.*
If

$$\lambda_{3N+1} > (1 - m_{2N+1}/m_N)/m_{3N+1}, \;\; \lambda_{k_2} > (1 - m_{k_2'}/m_{k_2''})/m_{k_2} \tag{3.28}$$

$k_2'' = \overline{N, 2N}$, $k_2' = \overline{2N+1, 3N}$, $k_2 = \overline{3N+1, 4N}$, *where for each* k_2 *it corresponds* k_2' *and* k_2''.

Then the queue length process $Q(\cdot)$ *is positive recurrent.*

In this case, when the higher priority classes are being served, the lower priority ones cannot be served, (class 1 cannot move to class 2 and 2 cannot move to 3,..., for further service, and vice versa. So, these later form a virtual stations. Therefore, these later, should not exceed one for the network to be stable. Now, let us consider the modified network. The additional classes 2N+1 and $\overline{3N+1, 4N}$ act as regulators that regulate the traffics. When the workloads of classes 3N+1 and $\overline{3N+2, 4N}$ are light, much service capacity of stations N+1,..., 2N are left to classes $\overline{2N+1, 3N}$ respectively and hence these later do not hold back the traffics to avoid building up of job queues at higher priority classes of the original network. Thus, the virtual stations effect prevails and the network is still unstable. However, when the workloads of classes $\overline{3N+1, 4N}$ are heavy enough such that the condition (3.28) holds, the service for lower priority classes $\overline{2N+1, 3N}$ is in effect slowed down and the traffics to the higher priority classes N and $\overline{N+2, 2N}$ are held back. Finally, the virtual stations effect is avoided and the modified network is thus stabilized.

Then, following the same steps given in theorem 3.3.1, it is not difficult to prove that there exists a time $\tau_1 \geq 0$ such that

$$\overline{Q}_{k_2}(t) = 0, \;\; k_2 = \overline{3N+1, 4N}, \;\; \text{for any } \; t \geq \tau_1. \tag{3.29}$$

after that, we prove that there exists a time $\tau_2 \geq \tau_1$ such that

$$\overline{Q}_N(t) = \overline{Q}_{k_2''}(t) = 0, \;\; k_2'' = \overline{N+2, 2N} \;\; \text{for any } \; t \geq \tau_2. \tag{3.30}$$

and finally, we prove that there exists a time $\tau \geq \tau_2 (\geq 0)$ such that

$$\overline{Q}_{k_4'}(t) = 0 \quad k_4' = \text{lower priority job class at stations } i = \overline{1, 2N}, \;\; \text{for } \; t \geq \tau \tag{3.31}$$

3.4 Conclusion

Multiclass queueing networks are effective tools for modelling many industrial settings. One setting for which the model is particularly attractive is the production flow within semiconductor manufacturing facilities.

In this paper we have studied the stabilization of N-stations queueing networks using its corresponding fluid network. The resulting model, fluid queueing networks with additional stations depending on the service priority and on the number of stations in the network are formally presented in Section 3. Beyond the presentation of our modified network models "fluid networks with additional stations", the primary concern of the paper is the stability of such networks. Nevertheless, stability of the artificial fluid model implies stability of the original network (see theorems 3.3.1 and 3.3.2).

Bibliographie

[1] Bramson, M, Stability of two families of queueing networks and a discussion of fluid limits, *Queueing Systems Theory and Applications.*, **23** (1998), 7-31.

[2] Chen, H, Fluid approximations and stability of multiclass queueing networks : Workconserving discipline, *Annals of Applied Probability.*, **5** (1995), 637-655.

[3] Chen, H. and D.D. Yao, *Fundamentals of Queueing Networks : Performance, Asymptotics and Optimization.*, Springer-Verlag New York, Inc.(2001)

[4] Chen, H. and Ye H.Q, Piecewise linear Lyapunov function for the stability of priority multiclass queueing networks, *IEEE Transactions on Automatic Control.*, **47**(4)(2002), 564-575.

[5] Chen, H. and H. Zhang, Stability of multiclass queueing networks under priority service disciplines, *Operations Research.*, **48**(2000), 26-37.

[6] Dai, J.G, On positive Harris recurrence of multiclass queueing networks : a unified approach via fluid models, *Annals of Applied Probability.*, **5**(1995), 49-77.

[7] Dai, J.G. A fluid-limit model criterion for instability of multiclass queueing networks, *Annals of Applied Probability.*, **6** (1996), 751-757.

[8] Dai, J.G. and Meyn, S.P, Stability and Convergence of moments for multiclass queueing networks via fluid models, *IEEE Transactions on Automatic Control.*, **40**(1995), 1899- 1904.

[9] Dai, J. G. and J. H. Vande Vate, Global Stability of Two-Station Queueing Networks. *Proceedings of Workshop on Stochastic Networks : Stability and Rare Events, Editors : Paul Glasserman, Karl Sigman and David Yao, Springer-Verlag, Columbia University, New York.*, (1996), 1-26.

[10] Heng-Qing Ye, A paradox for admission control of multiclass queueing network with differentiated service, *J. Appl. Probab.*, **44**(2) (2007), 321-331.

[11] Kumar, P.R. and T.I. Seidman, Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems, *IEEE Transactions on Automatic Control.*, **35**(1990), 289-298.

[12] Meyn, S, Transience of multiclass queueing networks via fluid limit models, *Annals of Applied Probability.*, **5**(1995), 946-957.

[13] Puhalskii, A. and Rybko, A.N, Non-ergodicity of queueing networks under nonstability of their fluid models, *Problems of information transmission.*, **36**(1)(2000), 26-48.

[14] Rybko, A.N. and Stolyar, A.L, Ergodicity of stochastic processed describing the operations of open queueing networks. *Problemy Peredachi Informatsii.*, **28**(1992), 2-26.

[15] Stolyar, A.L, On the stability of multiclass queueing network : a relaxed sufficient condition via limiting fluid processes, *Markov Process and Related Fields.*, **1**(4)(1995), 491-512.

Chapitre 4

A note on an $M/M/s$ queueing system with two reconnect and two redial orbits

Ce chapitre a fait l'objet d'une publication dans le journal Applications and Applied Mathematics : An International Journal (AAM)., Volume, 10(1), 1-12, 2015.

A note on an $M/M/s$ queueing system with two reconnect and two redial orbits

Amina Angelika BOUCHENTOUF[1], Hanane SAKHI[2]

[1] Department of Mathematics, Djillali Liabes university of Sidi Bel Abbes,
B. P. 89, Sidi Bel Abbes 22000, Algeria
E-mail :bouchentouf_amina@yahoo.fr

[2] Department of Mathematics, Sciences and Technology University of Oran : Mohamed Boudiaf, USTOMB.
B. P. 1505 EL-M'NAOUAR- Oran, Algeria.
E-mail :sakhi.hanane@yahoo.fr

Abstract.

A queueing system with two reconnect orbits, two redial (retrial) orbits, s servers and two independent Poisson streams of customers is considered. An arriving customer of type i, $i = 1, 2$ is handled by an available server, if there is any; otherwise, he waits in an infinite buffer queue. A waiting customer of type i who did not get connected to a server will lose his patience and abandon after an exponentially distributed amount of time, the abandoned one may leave the system (loss customer) or move into one of the redial orbits, from which he makes a new attempts to reach the primary queue, and when a customer finishes his conversation with a server, he may comeback to the system, to one of the reconnect orbits where he will be waiting for another service. In this paper, a fluid model is used to derive first order approximation for the number of customers in the redials and reconnect orbits in the heavy traffic. The fluid limit of such a model is the unique solution to a system of three differential equations.

Keywords : Queueing system, call center, reconnect queue, redial queue, retrial, fluid limit, abandonment, Bernoulli feedback.

2000 MSC No : Primary 60K25; Secondary 68M20; Thirdly 90B22

4.1 Introduction

In queueing theory, such a mechanism in which ejected (or rejected) customers return at random intervals until they receive service is called a retrial queue. Retrial queues have an important application in a wide variety of fields, they are also widespread in the evaluation and design of computer networks, in telecommunications, computer networks, and wireless networks. A retrial queue is similar to any ordinary queueing system in that there is an arrival process and one or more servers. The elemental differences are firstly, the customers who enter during a down or busy period of the server or servers may reattempt service at some random time in the future, and secondly a waiting room, which is known as a primary queue, in the context of retrial queues is not mandatory. In place of the ordinary waiting room is a buffer called an orbit queue (*in our work we call it a redial orbit*) to which customers proceed after an

unsuccessful attempt at service, and from which they retry service according to a given probabilistic or deterministic policy. Owing to the utility and interesting mathematical properties of retrial queueing models, an ample literature on the subject has emerged over the past several decades. For a general survey of retrial queues and a summary of many results, the reader may refer to the works of (Falin,1990 ; Gharbi & Ioualalen,2006 ; Falin & Artalejo,1998 ; Ebrahimi,2006 ; Libman & Orda,2002 ; Walrand,1991 ; Choi & Kim,1998) and references therein.
A queueing system with two orbits and two exogenous streams of different type serves as a model for two competing job streams in a carrier sensing multiple access system, where the jobs, after a failed attempt to network access, wait in an orbit queue (Nain,1985 ; Szpankowski,1994). An example of carrier sensing multiple access system is a local area computer network with bus architecture. The two types of customers can be interpreted as customers with different priority requirements.

A two-class retrial system with a single- server, no waiting room, batch arrivals and classical retrial scheme was introduced and analyzed in (Kulkarni,1986). Then, in (Falin, 1988) author extended the analysis of the model in (Kulkarni,1986) to the multi-class setting with arbitrary number of classes. In (Grishechkin,1992) author has established equivalence between the multi-class batch arrival retrial queues with classical retrial policy and branching processes with immigration. In (Mouzoukis & Langaris,1996) a non-preemptive priority mechanism was added to the model of (Falin, 1988, Kulkarni,1986). In (Langaris & Dimitriou, 2010) authors have considered a multi-class retrial system where retrial classes are associated with different phases of service.

Retrial queueing model $MMAP/M_2/1$ with two retrial orbits was studied in (Avrachenkov et al.,2010), authors considered a retrial single-server queueing model with two types of customers. When the server is occupied, the arrival customer moves to the orbit depending on the type of the customer, one orbit is infinite while the second one is a finite. Joint distribution of the number of customers in the orbits and some performance measures are computed. In (Bouchentouf & Belarbi,2013) authors considered two retrial queueing system with balking and feedback, the joint generating function of the number of busy server and the queue length was found by solving Kummer differential equation, and by the method of series solution.

At the present time, call centers are becoming an important way of communication with the customer. Therefore, the response-time performance of call centers is crucial for the customer satisfaction. Then, making the right staffing decisions is fundamental to the costs and the performances of call centers. Numerous models have been developed in order to decide on the right number of servers, see (Gans et al.,2002), (Halfin & Whitt,1981), and the references therein.

So, considering customer redial behaviors in call centers is quite significant (Gans et al.,2002), (Aguir et al.,2008), (Sze,1984) and reference therein.
In addition to redials (retrial queue), Reconnect is another important aspect, these two latter should be considered and modeled, see (Gans et al.,2002), (Ding, Van der Mei & Koole,2013). Appropriately, we can say that neglecting the impact of redials and reconnects will lead to either overstaffing or understating. In case of overstaffing the performance of the call center will be good, but at needlessly high costs. and

in case of understating, the performance of the call center will be degraded, and thus, it may lead to customer dissatisfaction.

When the system is heavily loaded, it would lead to bad service levels. However, for large call centers, especially during the busy hours when the inbound volume is quite large, it is possible that the target service levels can be met even in heavy traffic, for more details see (Garnett et al.,2002) and (Borst et al.,2004).

Fluid models for call centers have been extensively studied, (Whitt,2006), (Mandelbaum et al.,2002). In (Mandelbaum et al.,1999) the fluid and the diffusion approximation for time varying multiserver queue with abandonment and retrials were used, it was shown that the fluid and the diffusion approximation can both be obtained by solving sets of non-linear differential equations. In (Mandelbaum et al.,1998) more general theoretical results for the fluid and diffusion approximation for Markovian service networks was given. In (Aguir et al.,2004) authors extended the model by allowing customer balking behavior. Fluid models have also been applied in delay announcement of customers in call centers (Ibrahim & Whit,2009; Ibrahim & Whit,2011).

And recently, in (Ding et al., 2013) authors study call centers with one redial and one orbit, using fluid limit they calculate the expected total arrival rate, which is then given as an input to the Erlang. A model for the purpose of calculating service levels and abandonment rates. The performance of such a procedure is validated in the case of single intervals as well as multiple intervals with changing parameters.

In the present paper, an analysis of $M/M/s$ queueing model; a call center with two reconnect, two redials orbits and two exogenous streams of different types is carried out. The goal of this work is to present a type of call center where the emphasis of reconnect and redial orbits is crucial in any telecommunication system.

Now, let us outline the structure of the paper. After the introduction in section 2, we describe the queueing model. In section 3, we propose a fluid model, which is a deterministic analogue of the stochastic model. We prove that the original stochastic model converges to the fluid model under a proper scaling. So, in roughly, we use a fluid model to derive first order approximations for the number of customers in the redial and reconnect orbits in the heavy traffic, the fluid limit of such a model is the unique solution to a system of three differential equations.

4.2 The model

Consider a queueing model with two reconnect and two redial orbits and s servers (figure 4.1). Two independent Poisson streams of customers flow into a common infinite buffer queue. An arriving customer of type i, $i = 1, 2$ is handled by an available server, if there is any; otherwise, he waits in an infinite buffer queue. The customers are handled in the order of arrival (FIFO), The required service time of each customer is independent of its type.

A waiting customers of type i who did not get connected to a server will lose his patience (impatient customer) and abandon after an exponentially distributed amount of time ψ. We assume that $\mathbb{E}(\psi) =$

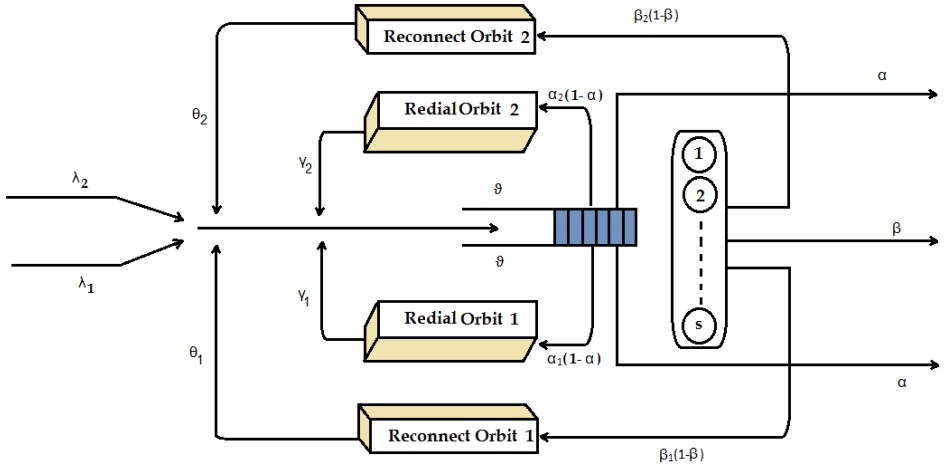

FIGURE 4.1 – A queueing model with two redial and two reconnect orbits

$\frac{1}{\vartheta} < \infty$, where ϑ is the abandonment rate. The abandoned customer either leaves definitively the system with probability α and he is considered as a lost customer or decides to stay requesting service with probability $(1-\alpha)$. So, if the customer decides to don't leave definitively the system, he will route to one of a separate type-i redial queue that attempts to re-dispatch its jobs with probability $(1-\alpha)\alpha_i$, $(i = 1, 2)$ $(\alpha_1 + \alpha_2 = 1)$. Which means that the customer have to choose between two redial orbits (for instance, customer of type 2 may be routed to redial orbit 1 with probability $(1-\alpha)\alpha_1$, or to redial orbit 2 with probability $(1-\alpha)\alpha_2$,), the choice is random and does not depend on the threshold of the redial orbits or on the type of the customer. Let's note that a redial orbit is an ordinary waiting room where a customer coming from the primary queue will be waiting there and then will redial after an exponentially distributed amount of time ω_i, (i=1,2) with $\mathbb{E}(\omega_i) = \gamma_i < \infty$. We assume that the service time of type-i customer has an exponential distribution with mean $\frac{1}{\mu}$. We assume that service times are independents, and the required service time of each customer is independent of its type. After the customer has been served, this later may leave the system with probability β or decide to comeback to the system for another service with probability $(1-\beta)$. The type-i customer will enter one of the type-i reconnect queue with probability $(1-\beta)\beta_i$, with $(\beta_1 + \beta_2 = 1)$. As, it has been mentioned above (in the case of redial orbits), let's note that the customer after getting service may comeback to the system, choosing between two reconnect orbits (for instance, customer of type 2 may be routed to reconnect orbit 1 with probability $(1-\beta)\beta_1$, or to reconnect orbit 2 with probability $(1-\beta)\beta_2$,), the choice is random and does not depend on the threshold of the retrial orbits or on the type of the customer. The reconnect orbit is an ordinary waiting room where a customer coming back from the last service requiring for another service will be waiting there and then will reconnect after an exponentially distributed time φ_i, with $\mathbb{E}(\varphi_i) = \theta_i < \infty$. So an abandoned customer of type i have to choose between one of the two retrial orbits, and the one who decide to comeback for another service have a choice between one of the two reconnect orbits, this means that a customer of type 1 (resp. of type 2) is not directed automatically to the redial or reconnect

orbits of type 1 (resp. of type 2).
We assume that α and β do not depend on customers' experiences in the system. These experiences include holding time, waiting time and the number of times that customers have already tried to get a service. This creates a system with five dependent queues. Such a queueing system serves as a model for two competing job streams in a carrier sensing multiple a access system.

4.3 Fluid approximation

In this section, we calculate $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ and $\mathbb{E}(Z_O)(t)$, where $Z_Q(t)$ is the number of customers in the queue plus the number of customers in service at time t, $\mathbb{E}(Z_R)(t) = \mathbb{E}(Z_{R_1})(t) + \mathbb{E}(Z_{R_2})(t)$ is the number of customers in the redial queues 1 and 2 at time t, and $\mathbb{E}(Z_O)(t) = \mathbb{E}(Z_{O_1})(t) + \mathbb{E}(Z_{O_2})(t)$ is the number of customers in reconnect orbits 1 and 2 at time t.

So, let

$$\mathbb{E}\Delta(t) \quad = \quad \lambda_1(t) + \lambda_2(t) + (\gamma_1 + \gamma_1)\mathbb{E}(Z_R(t)) + (\theta_1 + \theta_2)\mathbb{E}(Z_O(t)); \tag{4.1}$$

where $\Delta(t)$ stands for the expected number of arrivals up to time t, which is a stochastic process, $\lambda_1(t) + \lambda_2(t)$ stands for the external arrival rate at time t, $(\gamma_1 + \gamma_1)\mathbb{E}(Z_R(t))$ and $(\theta_1 + \theta_2)\mathbb{E}(Z_O(t))$ stand for the mean arrival rate due to redials and reconnects at time t, respectively. Therefore, once $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ and $\mathbb{E}(Z_O)(t)$, are known, $\mathbb{E}\Delta(t)$ can be obtained by equation (4.1). Note that $Z_Q(t)$ does not appear in Equation (4.1), but we will see later that $Z_R(t)$ and $Z_O(t)$ would depend on $Z_Q(t)$.

In fact, it is easy to prove that our stochastic process $\{Z(t); t \geq 0\}$, which is defined by

$$Z(t) := (Z_Q(t), Z_{R_1}(t), Z_{R_2}(t), Z_{O_1}(t), Z_{O_2}(t))^T; \tag{4.2}$$

is a 5-dimensional Markov process, The state space of this Markov process is $\mathbb{Z}_+^5$. This stochastic process can be reduced to 3-dimensional Markov process

$$(Z_Q(t); Z_R(t); Z_O(t))^T,$$

with $Z_R(t) = Z_{R_1}(t) + Z_{R_2}(t)$ and $Z_O(t) = Z_{O_1}(t) + Z_{O_2}(t)$. Then along all the paper we will focus on the study of 3-dimensional Markov process, because we think that is not necessary to study each of the reconnect and redial orbits separately.
And since it is a Markov process, we can truncate the system at certain large state, and numerically obtain the steady state distribution of $Z(t)$ by solving global balance equations. But, for the model we consider, it is difficult to formulate and solve the global balance equations, and their solution offers no insight about the system. Therefore, for the convenience of practical usage, we will not consider solving this Markov process, but other approximation methods.

4.3.1 Fluid limit

In this subsection, we present the fluid model, which we show to arise as the limit under a proper scaling of the stochastic model presented in Figure 1. Consider a single interval with the external arrival

rates remaining constant during this interval ($\lambda_1(t) + \lambda_2(t) = \lambda_1 + \lambda_2,\ t \geq 0$). The following flow conservation equations hold for this stochastic model

$$Z_Q(t) = Z_Q(0) + \Delta_{\lambda_1}(t) + \Delta_{\lambda_2}(t) + D_R(t) + D_O(t) - D_s(t) - D_a(t). \tag{4.3}$$

With

$$D_{s_1}(t) + D_{s_2}(t) = D_s(t) \quad \text{and} \quad D_{a_1}(t) + D_{a_2}(t) = D_a(t).$$

$$Z_R(t) = Z_R(0) + \sum_{j=1}^{D_a(t)} B_j(1-\alpha) - D_R(t). \tag{4.4}$$

$$Z_O(t) = Z_O(0) + \sum_{j=1}^{D_s(t)} B_j(1-\beta) - D_O(t). \tag{4.5}$$

Where $\Delta_{\lambda_1}(t) + \Delta_{\lambda_2}(t)$ is the number of external arrivals of type 1 and 2 during time interval $[0;t)$. $\Delta_{\lambda_1}(\cdot)$, and $\Delta_{\lambda_2}(\cdot)$ are a Poisson processes of rates λ_1, and λ_2 respectively. $D_R(t) = D_{R_1}(t) + D_{R_2}(t)$ is the number of redials during $[0;t)$, $D_O(t) = D_{O_1}(t) + D_{O_2}(t)$; the number of reconnects during $[0;t)$, $D_s(t) = D_{s_1}(t) + D_{s_2}(t)$; number of served customers of type 1 and 2, during $[0;t)$, and $D_a(t) = D_{a_1}(t) + D_{a_2}(t)$ the number of abandoned customers of type 1 and 2 during $[0;t)$. Finally $B_j(1-\alpha)$ is a Bernoulli random variable with success probability $1-\alpha$, $j = 1, ..., D_{a_i}(t)$, such that

$$B_j(1-\alpha) = \begin{cases} 1, & \text{if the } jth \text{ abandoned customer decides to stay in the system and then} \\ & \text{enters one of the redial orbits;} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, for given $D_a(t)$, we have $\sum_{j=1}^{D_a(t)} B_j(1-\alpha) \rightsquigarrow Bin(D_a(t), 1-\alpha)$. And for a given $D_s(t)$, we have

$\sum_{j=1}^{D_s(t)} B_j(1-\beta) \rightsquigarrow Bin(D_s(t), 1-\beta)$, ($Bin(\cdot,\cdot)$ is a Binomial distribution)

Let Δ_i be independent Poisson processes of rate 1.

So,

$$D_s(t) = \Delta_1\left(\int_0^t \mu \min\{s, Z_Q(u)\} du\right), \tag{4.6}$$

$$D_a(t) = \Delta_2\left(\int_0^t \vartheta (Z_Q(u) - s)^+ du\right), \tag{4.7}$$

$$D_R(t) = \Delta_3\left(\int_0^t (\gamma_1 + \gamma_2) Z_R(u) du\right), \tag{4.8}$$

$$D_O(t) = \Delta_4\left(\int_0^t (\theta_1 + \theta_2) Z_O(u) du\right). \tag{4.9}$$

We do not need to give the proof of these statements, it is sufficient to refer to (Pang et al., 2007).

To introduce the fluid limit, we consider a sequence of models as in Figure 1 such that, in the n-th model,

the external arrival rates are $\lambda_1 n$, $\lambda_2 n$ and the number of servers is ns. We add the superscript "(n)" to all notations in the n-th model. Similarly to (4.3)-(4.5), we then have for the n$-th$ model :

$$Z_Q^{(n)}(t) = Z_Q^{(n)}(0) + \Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t) + D_R^{(n)}(t) + D_O^{(n)}(t) - D_s^{(n)}(t) - D_a^{(n)}(t), \tag{4.10}$$

$$Z_R^{(n)}(t) = Z_R^{(n)}(0) + \sum_{j=1}^{D_a^{(n)}(t)} B_j(1-\alpha) - D_R^{(n)}(t). \tag{4.11}$$

$$Z_O^{(n)}(t) = Z_O^{(n)}(0) + \sum_{j=1}^{D_s^{(n)}(t)} B_j(1-\beta) - D_O^{(n)}(t). \tag{4.12}$$

Now, dividing by n on both sides of equations (4.10)-(4.12),we have

$$\bar{Z}_Q^{(n)}(t) = \bar{Z}_Q^{(n)}(0) + G_Q^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_Q\left(\bar{Z}^{(n)}\right)(u)du, \tag{4.13}$$

$$\bar{Z}_R^{(n)}(t) = \bar{Z}_R^{(n)}(0) + G_R^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_R\left(\bar{Z}^{(n)}\right)(u)du, \tag{4.14}$$

$$\bar{Z}_O^{(n)}(t) = \bar{Z}_O^{(n)}(0) + G_O^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_O\left(\bar{Z}^{(n)}\right)(u)du, \tag{4.15}$$

Where

$$\begin{aligned} G_Q^{(n)}\left(\bar{Z}^{(n)}\right)(t) = & \left(\frac{\Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t)}{n} - (\lambda_1+\lambda_2)t\right) \\ & - \left(\bar{D}_s^{(n)}(t) - \int_0^t \mu \min\{s, \bar{Z}_Q^{(n)}(u)\}du\right) \\ & - \left(\bar{D_a}^{(n)}(t) - \int_0^t \vartheta(\bar{Z}_Q^{(n)}(u) - s)^+ du\right) \\ & + \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)du\right) \\ & + \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1+\theta_2)\bar{Z}_O^{(n)}(u)du\right) \end{aligned} \tag{4.16}$$

$$\begin{aligned} G_R^{(n)}\left(\bar{Z}^{(n)}\right)(t) = & \left(\sum_{j=1}^{n\bar{D_a}^{(n)}(t)} B_j(1-\alpha)/n - \int_0^t (1-\alpha)\vartheta\left(\bar{Z}_Q^{(n)}(u) - s\right)^+ du\right) \\ & - \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)du\right) \end{aligned} \tag{4.17}$$

$$\begin{aligned} G_O^{(n)}\left(\bar{Z}^{(n)}\right)(t) = & \left(\sum_{j=1}^{n\bar{D_s}^{(n)}(t)} B_j(1-\beta)/n - \int_0^t (1-\beta)\mu \min\{s, \bar{Z}_Q^{(n)}(u)\}du\right) \\ & - \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1+\theta_2)\bar{Z}_O^{(n)}(u)du\right). \end{aligned} \tag{4.18}$$

And

$$\bar{D_s}^{(n)}(t) = \Delta_1 \left(n \int_0^t \mu \min\{s, \bar{Z}_Q^{(n)}(u)\} du \right) /n \tag{4.19}$$

$$\bar{D_a}^{(n)}(t) = \Delta_2 \left(n \int_0^t \vartheta (\bar{Z}_Q^{(n)}(u) - s)^+ du \right) /n \tag{4.20}$$

$$\bar{D}_R^{(n)}(t) = \Delta_3 \left(n \int_0^t (\gamma_1 + \gamma_2) \bar{Z}_R^{(n)}(u) du \right) /n \tag{4.21}$$

$$\bar{D}_O^{(n)}(t) = \Delta_4 \left(n \int_0^t (\theta_1 + \theta_2) \bar{Z}_O^{(n)}(u) du \right) /n \tag{4.22}$$

And

$$\begin{aligned} \int_0^t H_Q \left(\bar{Z}^{(n)} \right)(u) du = \quad & \int_0^t \lambda_1 + \lambda_2 + (\gamma_1 + \gamma_2) \bar{Z}_R^{(n)}(u) + (\theta_1 + \theta_2) \bar{Z}_O^{(n)}(u) \\ & - \mu \min\{s, \bar{Z}_Q^{(n)}(u)\} - \vartheta \left(\bar{Z}_Q^{(n)}(u) - s \right)^+ du \end{aligned} \tag{4.23}$$

$$\int_0^t H_R \left(\bar{Z}^{(n)} \right)(u) du = \int_0^t (1 - \alpha) \vartheta \left(\bar{Z}_Q^{(n)}(u) - s \right)^+ - (\gamma_1 + \gamma_2) \bar{Z}_R^{(n)}(u) du \tag{4.24}$$

$$\int_0^t H_O \left(\bar{Z}^{(n)} \right)(u) du = \int_0^t (1 - \beta) \mu \min\{s, \bar{Z}_Q^{(n)}(u)\} - (\theta_1 + \theta_2) \bar{Z}_O^{(n)}(u) du \tag{4.25}$$

For the convenience of notation, we rewrite equations (4.13)-(4.15) in the vector form

$$\bar{Z}^{(n)}(t) = \bar{Z}^{(n)}(0) + G^{(n)} \left(\bar{Z}^{(n)} \right)(t) + \int_0^t H_Q(\bar{Z}^{(n)})(u) du \tag{4.26}$$

where

$$G^{(n)} \left(\bar{Z}^{(n)} \right)(t) = (G_Q^{(n)} \left(\bar{Z}^{(n)} \right)(t), G_R^{(n)} \left(\bar{Z}^{(n)} \right)(t), G_O^{(n)} \left(\bar{Z}^{(n)} \right)(t))^T$$

$$H^{(n)} \left(\bar{Z}^{(n)} \right)(u) = (H_Q^{(n)} \left(\bar{Z}^{(n)} \right)(u), H_R^{(n)} \left(\bar{Z}^{(n)} \right)(u), H_O^{(n)} \left(\bar{Z}^{(n)})(u) \right)^T$$

Now, let us define the fluid scaled process

$$\bar{Z}^n(t) = (\bar{Z}_Q^{(n)}(t), \bar{Z}_R^{(n)}(t), \bar{Z}_O^{(n)}(t))^T$$

Where

$$\bar{Z}_Q^{(n)}(t) = \frac{Z_Q^{(n)}(t)}{n}, \quad \bar{Z}_R^{(n)}(t) = \frac{Z_R^{(n)}(t)}{n}, \quad \bar{Z}_O^{(n)}(t) = \frac{Z_O^{(n)}(t)}{n}.$$

Let $D([0, \infty), \mathbb{R}^3)$ be the space of right continuous functions with left limits in $\mathbb{R}^3$ having the domain $[0, \infty)$. We endow $D([0, \infty), \mathbb{R}^3)$ with the usual Skorokhod J_1 topology. Assume that $\{X^{(n)}\}_{n=1}^{\infty}$ is a sequence of stochastic processes, then $X^{(n)} \to^d x$ means that $X^{(n)}$ converge weakly to stochastic process x.

Definition 4.3.1 *If there exists a limit in distribution for the scaled process* $\{\bar{Z}^{(n)}(\cdot)\}_{n=1}^{\infty}$, *i.e.* $\bar{Z}^{(n)}(\cdot) \to^d z(\cdot)$, *then* $z(\cdot)$ *is called the fluid limit of the original stochastic model.*

Now, let us define the fluid limit of the original stochastic model ; so if

$$H_0 : \ (\overline{Z}_Q^{(n)}(0), \overline{Z}_R^{(n)}(0), \overline{Z}_O^{(n)}(0)) \to^d (z_Q(0), z_R(0), z_O(0)) \ \text{ as } \ n \to \infty,$$

holds, then the fluid limit of the original stochastic model is the unique solution to the following system of equations

$$\begin{aligned} z_Q(t) = z_Q(0) + (\lambda_1 + \lambda_2)t + (\gamma_1 + \gamma_2)\int_0^t z_R(u)du + (\theta_1 + \theta_2)\int_0^t z_O(u)du - \\ \mu \int_0^t \min\{s, z_Q(u)\}du - \vartheta \int_0^t (z_Q(u) - s)^+ du \end{aligned} \tag{4.27}$$

$$z_R(t) = z_R(0) + (1-\alpha)\vartheta \int_0^t (z_Q(u) - s)^+ du - (\gamma_1 + \gamma_2)\int_0^t z_R(u)du. \tag{4.28}$$

$$z_O(t) = z_O(0) + (1-\beta)\mu \int_0^t \min\{s, z_Q(u)\}du - (\theta_1 + \theta_2)\int_0^t z_O(u)du. \tag{4.29}$$

To demonstrate this latter, we have to refer to (Ding et al.,2013), the proof will be given in very similar way.

Now, in the next section, by using equations (4.27)-(4.29) given above we derive the fluid limit in stationarity.

4.4 The stationarity of the model

In this section we develop conditions under which $z(t)$ is constant. By differentiating Equations (4.27)-(4.29), we obtain

$$\lambda_1 + \lambda_2 = \mu \min\{s, z_Q(\infty)\}) + \vartheta(z_Q(\infty) - s)^+ - (\gamma_1 + \gamma_2)z_R(\infty) - (\theta_1 + \theta_2)z_O(\infty). \tag{4.30}$$

$$(\gamma_1 + \gamma_2)z_R(\infty) = (1-\alpha)\vartheta(z_Q(\infty) - s)^+. \tag{4.31}$$

$$(\theta_1 + \theta_2)z_O(\infty) = (1-\beta)\mu \ \min\{s, z_Q(\infty)\}. \tag{4.32}$$

Where $z_Q(\infty) = \lim\limits_{t \longrightarrow \infty} z_Q(t)$, $z_R(\infty) = \lim\limits_{t \longrightarrow \infty} z_R(t)$, $z_O(\infty) = \lim\limits_{t \longrightarrow \infty} z_O(t)$. Equations (4.30)-(4.32) can be easily solved with respect to $z_Q(\infty)$, $z_R(\infty)$ and $z_O(\infty)$, yielding

$$z_Q(\infty) = \begin{cases} \frac{\lambda_1 + \lambda_2}{\mu\beta} & \text{if } \ \frac{\lambda_1 + \lambda_2}{s\mu} < \beta \\ \\ \frac{\lambda_1 + \lambda_2 - \beta\mu s}{\vartheta\alpha} + s & \text{otherwise} \end{cases} \tag{4.33}$$

$$z_R(\infty) = \begin{cases} 0 & \text{if } \frac{\lambda_1+\lambda_2}{s\mu} < \beta \\ \frac{(1-\alpha)\vartheta(z_Q(\infty)-s)}{\gamma_1+\gamma_2} & \text{otherwise.} \end{cases} \tag{4.34}$$

$$z_O(\infty) = \begin{cases} \frac{(1-\beta)\mu z_Q(\infty)}{\theta_1+\theta_2} & \text{if } \frac{\lambda_1+\lambda_2}{s\mu} < \beta \\ \frac{(1-\beta)\mu s}{\theta_1+\theta_2} & \text{otherwise} \end{cases} \tag{4.35}$$

4.4.1 Discussion on the results

At first, let us notice that $\frac{\lambda_1+\lambda_2}{s\mu}$ is the load of the system due to the external arrivals. $\frac{\lambda_1+\lambda_2}{\mu\beta}$ is the total load when there is no redials. Then let us notice that in expressions (4.33), (4.34) and (4.35), the value of $\frac{\lambda_1+\lambda_2}{\mu\beta}$ determines whether the fluid model is in heavy traffic or not. So, now if $\frac{\lambda_1+\lambda_2}{\mu\beta} < 1$, since the fluid limit is deterministic, we get $z_Q(\infty) < s$, and there is no customers in the two redial orbits which means that $z_R(\infty) = 0$ would hold, which means also that there is no abandonment at all in the fluid limit. In reality, due to the variability of the service duration and patience, abandonments would not be 0 though, but very small. Now, if $\frac{\lambda_1+\lambda_2}{\mu\beta} > 1$, by equation (4.33), we have $z_Q(\infty) > s$. Consequently, the fluid model indicates that there will be $z_Q(\infty) - s$ amount of customers waiting, each with rate ϑ, and customers will be routed the redial orbits with rate $(1-\alpha)\vartheta(z_Q(\infty) - s)$.

4.5 Conclusion

In our paper, a call center with two retrial orbits, two redial orbits and two exogenous streams of different types was studied ; we analyzed this system where customers can abandon, and the abandoned one may redial, and when a customer finishes service, he may reconnect.
In this work, a fluid model is used to derive first order approximations for the number of customers in the redial and reconnect orbits in the heavy traffic, the fluid limit of such a model is the unique solution to a system of three differential equations.

Acknowledgments

The authors are thankful to anonymous referees and the Editor-in-Chief Professor Aliakbar Montazer Haghighi for useful comments and suggestions towards the improvement of this paper.

REFERENCES

M.S. Aguir, F. Karaesmen, O.Z. Aksin, and F. Chauvet.(2004). The impact of retrials on call center performance. *OR Spectrum*, 26(3) :353-376.

M.S. Aguir, O.Z. Akşin, F. Karaesmen, and Y. Dallery.(2008). On the interaction between retrials and sizing of call centers. *European Journal of Operational Research*, 191(2) :398-408.

Avrachenkov, K., Dudin, A., and Klimenok, V. (2010). Queueing Model $MMAP/M_2/1$ with Two Orbits, *Lecture Notes in Computer Science*, 6235 :107-118.

S. Borst, A. Mandelbaum, and M. Reiman.(2004). Dimensioning large call centers. *Operations Research*, 52(1) :17-34.

A.A. Bouchentouf & F. Belarbi.(2013). Performance evaluation of two Markovian retrial queueing model with balking and feedback, *Acta Univ. Sapientiae, Mathematica,* 5(2) :132-146

B. D. Choi and Y. C. Kim.(1998). The $M/M/c$ Retrial Queue with Geometric Loss and Feedback, *Computers Math. Applic*, 36 (6) :41-52.

S. Ding, R.D. van der Mei, and G. Koole. *A method for estimation of redial and reconnect probabilities in call centers.* In Proceedings of the 2013 Winter Simulation Conference, Winter Simulation Conference, 2013.

S. Ding , M. Remerova, Rob van der Mei, B. Zwart, *Fluid approximation of a Call Center model with redials and reconnects*, Cornell University Library,arXiv :1311.6248v1 [Math.PR], 25 November 2013.

N. Ebrahimi.(2006). System reliability based on system wear. *Stochastic Models*, 22(1) :21-36.

Falin, G.I.(1988). On a multiclass batch arrival retrial queue. *Adv. Appl. Probab.* 20 :483-487.

G. Falin.(1990). A survey of retrial queues, *Queueing Systems : Theory and Applications*, 7(2) :127-167.

G. I. Falin and J. R. Artalejo.(1998). An infinite source retrial queue. *European Journal of Operational Research*, 108(2) :409-424.

N Gans, G Koole, and A Mandelbaum. (2002). *Telephone calls centers : a tutorial and literature review.*European Jour. Oper. Res.

N. Gharbi and M. Ioualalen.(2006). GSPN analysis of retrial systems with server break- downs and repairs. *Applied Mathematics and Computation*, 174(2) :1151-1168.

O. Garnett, A. Mandelbaum, and M. Reiman.(2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3) :208-227.

Grishechkin, S.A.(1992). Multiclass batch arrival retrial queues analyzed as branching processes with immigration. *Queueing Syst*, 11 :395-418.

S. Halfin and W. Whitt.(1981). Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3) :567-588.

R. Ibrahim and W. Whitt.(2009). Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management*, 11(3) :397-415.

R. Ibrahim and W. Whitt.(2011). Wait-time predictors for customer service systems with time-varying demand and capacity. *Operations Research*, 59(5) :1106-1118.

Kulkarni, V.G.(1986). Expected waiting times in a multiclass batch arrival retrial queue. *J. Appl. Probab*, 23 :144-154.

Langaris, C., Dimitriou, I.(2010). A queueing system with n-phases of service and $(n-1)-$types of retrial customers. *Eur. J. Oper. Res*, 205 :638-649.

L. Libman and A. Orda.(2002). Optimal retrial and timeout strategies for accessing network resources. *IEEE/ACM Transactions on Networking*, 10(4) :551-564.

A. Mandelbaum, W. Massey.(1998). and M. Reiman. Strong approximations for markovian service

networks. *Queueing Systems*, 30(1-2) :149-201 .

A. Mandelbaum, W. Massey, M. Reiman, and B. Rider.(1999). Time varying multiserver queues with abandonment and retrials. *In Proceedings of the 16th International Teletraffic Conference*, 4 :4-7.

A. Mandelbaum, W. Massey, M. Reiman, A. Stolyar, and B. Rider.(2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. *Telecommunication Systems*, 21(2-4) :149-171.

Moutzoukis, E., Langaris, C.(1996). Non-preemptive priorities and vacations in a multiclass retrial queueing system. *Stoch Models*, 12(3) :455-472.

Nain, P. : Analysis of a two-node Aloha network with infinite capacity buffers. In : Hasegawa, T., Takagi, H., Takahashi, Y.(1985). (eds.) Proc. Int. *Seminar on Computer Networking and Performance Evaluation*. Tokyo, Japan,18-20.

G. Pang, R. Talreja, and W. Whitt.(2007) Martingale proofs of many-server heavy-traffic limits for markovian queues. *Probability Surveys*, 4 :193-267.

A. Papoulis, *Probability Random Variables and Stochastic Processes*, 2nd ed., McGraw-Hill,1983.

D. Sze.(1984). Or practice - a queueing model for telephone operator staffing. *Operations Re- search*, 32(2) :229-249.

Szpankowski, W.(1994). Stability conditions for some multiqueue distributed systems : buffered random access systems. *Adv. Appl. Probab*, 26 :498-515.

J, Walrand, *Communication Networks : A First Course,* The Aksen Associates Se- ries in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA. 1991.

W. Whitt.(2006). Fluid models for multiserver queues with abandonments. *Operations Research*, 54(1) :37-54.

Conclusion Générale

Dans cette thèse nous nous sommes intéréssés aux systèmes de files d'attente multiclasses fluides.

► *Le chapitre un est une introduction pour les systèmes de files d'attente multiclasse, et basé sur le fait que la stabilité d'un réseau de file d'attente multiclasse est implicite par la stabilité du réseau fluide associé, nous avons ainsi présenté les propriétés de réseaux fluides sous diverses disciplines.*

► *Dans les trois derniers chapitres, nous avons réalisé*

⁎ *L'analyse de la stabilité d'un système de files d'attente multiclasses avec priorité, dont la région de stabilité globale de ce réseau de files d'attente composé de N unités, $N \geq 3$ et N^2 classes de clients, en utilisant la fonction de Lyapunov linéaire par morceaux est déterminée.*

⁎ *L'étude de la stabilisation d'un réseau de files d'attente multiclasses avec N stations et $2N$ classes de clients en utilisant son réseau fluide correspondant ; un réseau de files d'attente fluide avec des stations supplémentaires en fonction de la priorité de service et le nombre de stations du réseau.*

⁎ *L'analyse d'un système de files d'attente $M/M/s$, avec des orbites réservées aux clients impatients dans le système dites "redial orbits" et d'autres réservées aux client revenant de l'éxtérieur pour d'autres services dites "reconnect orbits". Un modèle fluide est utilisé pour calculer l'approximation d'ordre premier pour le nombre de clients dans les orbites du système.*

Les résultats obtenus, constituent une base de départ importante dans le domaine de communication, d'informatique et de fabrication. Beaucoup d'efforts ont été dépensés pour dériver les conditions qui garantissent la stabilité des réseaux fluides sous diverses disciplines présentés dans cette thèse.

Comme perspectives, nous avons l'intention de mener des études similaire pour des modèles de files d'attente avec casacades et différents services de priorité.

Annexe

La récurrence au sens de Harris

Dans cette partie, nous présentons la base de l'analyse de la stabilité des réseaux de files d'attente multiclasses. Pour cela, nous considérons un processus Markov X avec des valeurs dans un espace métrique séparable et localement compact E. L'algébre de Borel induite par la métrique est notée $\mathbf{B}(E)$. En outre, soit P la fonction de transition associée à X telle que pour $A \in \mathbf{B}(E)$ nous avons,

$$P(t,x,A) = \mathbb{P}_x[X(t) \in A] \quad \textit{avec } x \textit{ est la valeur initiale.}$$

La mésure ν dans $(E, \mathbf{B}(E))$ est dite invariante par rapport à X si elle est σ- finie et nous avons,

$$\nu(A) = \int_E P(t,x,A)\nu(dx),$$

pour tout $A \in \mathbf{B}(E)$ et $t \geq 0$. Pour tout $A \in \mathbf{B}(E)$

$$\tau_A = \inf\{t \geq 0 : \quad X(t) \in A\},$$

est dit le temps d'atteinte de A, et pour $\delta > 0$ le temps d'atteinte de A avant δ est défini par

$$\tau_A(\delta) = \inf\{t \geq \delta : X(t) \in A\}.$$

Par le théorème de Sharpe (1988), le temps d'atteinte est un temps d'arrêt.
En outre pour $A \in \mathbf{B}(E)$, nous considérons le temps d'occupation η_A, décrivant le nombre de visites de X à A par,

$$\eta_A = \int_0^\infty 1_{\{X(t) \in A\} dt}.$$

Le processus de Markov est dit récurrent au sens de Harris s'il existe une mesure non nulle σ- finie telle que pour $\nu(A) > 0$ et $A \in \mathbf{B}(E)$ nous avons,

$$\mathbb{P}_x(\eta_A = \infty) = 1 \quad \text{pour tout } x \in E.$$

Getoor (1980) a montré que pour les processus récurrents au sens de Harris, il existe une mesure invariante

unique. Si la mésure invariante unique peut être normalisée à une mésure de probabilité, le processus de Markov est dit récurrent positif au sens de Harris.

Maintenant, nous présentons une caractérisation de la récurrence positive au sens de Harris qui est plus facile à appliquer. Supposons que a est une mesure de probabilité sur $(0,\infty)$ et nous considérons le processus de Markov X_a avec une fonction de transition,

$$T_a(x,A) = \int_0^\infty P(t,x,A)a(dt),$$

où $X \in E$ et $A \in \mathbf{B}(E)$. L'ensemble non-vide $A \in \mathbf{B}(E)$ $(A \neq \emptyset)$ est dit petit s'il existe une mésure μ non nulle sur $(E,\mathbf{B}(E))$ et une autre mésure de probabilité a sur $(0,\infty)$ telle que la fonction de transition $T_a(x,B)$ du processus de l'échantillon satisfait,

$$T_a(x,B) \geq \mu(B),$$

pour tout $x \in A$ et pour tout $B \in \mathbf{B}(E)$. Le petit ensemble A a la propriété suivante : tout ensemble $B \in \mathbf{B}(E)$ est accessible à partir de de tous les points $x \in A$ par rapport à la mésure μ. Le résultat suivant présente des conditions qui sont trés importants pour la récurrence au sens de Harris dans l'analyse des réseaux de file d'attente multiclasses.

Théorème 4.5.1 (Bramson (2008)) *Soit X un processus de Markov*

• *X est récurrent au sens de Harris si et seulement s'il existe un petit ensemble fermé A tel que pour tout $x \in E$ nous avons*

$$\mathbb{P}_x(\tau_A < \infty) = 1 \quad \textit{pour tout}, \quad x \in E.$$

• *Si x est récurrent au sens de Harris. Alors, X est récurrent positif au sens de Harris si et seulement si il existe un petit ensemble fermé A tel que pour certain $\delta > O$,*

$$\sup_{x\in A} \mathbb{E}_x[\tau_A(\delta)] < \infty.$$

BIBLIOGRAPHIE GENERALE

M.S. Aguir, F. Karaesmen, O.Z. Aksin, and F. Chauvet. (2004). The impact of retrials on call center performance. OR Spectrum, 26(3) : 353-376.

M.S. Aguir, O.Z. Akşin, F. Karaesmen, and Y. Dallery. (2008). On the interaction between retrials and sizing of call centers. European Journal of Operational Research, 191(2) : 398-408.

K. Avrachenkov, A. Dudin, and V. Klimenok. (2010). Queueing Model $MMAP/M_2/1$ with Two Orbits. Lecture Notes in Computer Science, 6235 : 107-118.

A. Bacciotti and L. Rosier. (2005). Liapunov functions and stability in control theory. 2nd ed. Communications and Control Engineering. Springer, Berlin Heidelberg.

D. Bertsimas, D.Gamarnik and J.N.Tsitsiklis. (1996). Stability conditions for multiclass fluid queueing networks. IEEE Trans. Automat. Control. 41, no. 11, 1618-1631.

D. Bertsimas, I. Ch. Paschalidis, and J. N. Tsitsiklis. (1994). Optimization of multiclass queueing networks : polyhedral and nonlinear characterizations of achievable performance. Annals of Applied Probability 4 : 43-75.

S. Borst, A. Mandelbaum, and M. Reiman. (2004). Dimensioning large call centers. Operations Research, 52(1) : 17-34.

D.D. Botvich and A.A. Zamyatin. (1992). Ergodicity of conservative communication networks. Rapport de recherche, 1772, INRIA.

A.A. Bouchentouf & F. Belarbi. (2013). Performance evaluation of two Markovian retrial queueing model with balking and feedback. Acta Univ. Sapientiae. Mathematica, 5(2) : 132-146

M. Bramson. (1994). Instability of FIFO queueing networks. Ann. Appl. Probab., 4(2) : 414-431.

M. Bramson. (1996). Convergence to equilibria for fluid models of FIFO queueing networks. Queueing Syst., 22(1-2) : 5-45.

M. Bramson. (1996). Convergence to equilibria for fluid models of head-ofthe- line proportional processor sharing queueing networks. Queueing Syst., 23(1-4) : 1-26.

M. Bramson. (1998). Stability of two families of queueing networks and a discussion of fluid limits. Queueing Systems Theory and Applications, 23, 7-31.

M. Bramson. (2008). Stability of queueing networks. Lecture Notes in Mathematics 1950, Springer, Berlin Heidelberg.

H. Chen. (1995). Fluid approximations and stability of multiclass queueing networks : Work-conserving disciplines. Ann. Appl. Probab., 5(3) : 637-665.

H. Chen and D.D. Yao. (2001). Fundamentals of Queueing Networks : Performance, Asymptotics and Optimization. Springer-Verlag New York, Inc.

H. Chen and H. Ye. (2002). Piecewise linear Lyapunov function for the stability of multiclass priority queueing networks. IEEE Trans. Autom. Control, 47(4) : 564-575.

H. Chen and H. Zhang. (1997). Stability of multiclass queueing networks under FIFO service discipline. Math. Oper. Res., 22(3) : 691-725.

H. Chen and H. Zhang. (2000). Stability of multiclass queueing networks under priority service disciplines. Oper. Res., 48(1) : 26-37.

B. D. Choi and Y. C. Kim.(1998). The $M/M/c$ Retrial Queue with Geometric Loss and Feedback. Computers Math. Applic, 36 (6) : 41-52.
J. G. Dai. (1995). On positive Harris recurrence of multiclass queueing networks : a unified approach via fluid limit models. Annals of Applied Probability 5 : 49-77.

J. G. Dai. (1995). On possitive Harris recurrence of multiclass queueing networks : a unified approach via fluid limit models. Mathematics and its applications, 71, 71-90.

J. Dai. (1996). A fluid limit model criterion for instability of multiclass queueing networks. Ann. Appl. Probab., 6(3) : 751-757.

J. Dai. (1999). Stability of fluid and stochastic processing networks. Ma- PhySto. Miscellanea. 9. Aarhus : Univ. of Aarhus.

J. Dai. (1995). Stability of open multiclass queueing networks via fluid models. In Kelly, F. and Williams, R. (Eds.), Stochastic Networks. Volume 71 of IMA Volumes in Mathematics and Its Applications, pp. 71-90, Springer New York.

J. Dai, J. J. Hasenbein, and B. Kim. (2007). Stability of join-the-shortestqueue networks. Queueing Syst., 57(4) : 129-145.

J. G. Dai, J. J. Hasenbein and J. Vande Vate.(1999). stability of a three-station fluid network. Queueing

Systems, 33, 293-325.

J. Dai, J. J. Hasenbein, and J. H. Vande Vate. (2004). Stability and instability of a two-station queueing network. Ann. Appl. Probab., 14(1) : 326-377.

J. Dai and S.P. Meyn. (1995). Stability and Convergence of moments for multiclass queueing networks via fluid models. IEEE Transactions on Automatic Control, 40, 1899- 1904.

J. G. Dai and J. Vande Vate. (1996). Global stability of two-station queueing networks. Lecture Notes in Statistics. Columbia University, New York, Springer-Verlag, 117, 1-26.

J. G. Dai and J. Vande Vate. (2000). The stability of two-station fluid networks. Operations Research, 48, 721-744.

J. G. Dai and G. Weiss. (1996). Stability and instability of fluid models for re-entrant lines. Mathematics of operations Research, 21, 115-134.

S. Ding, R.D. van der Mei, and G. Koole. (2013). A method for estimation of redial and reconnect probabilities in call centers. In Proceedings of the 2013 Winter Simulation Conference, Winter Simulation Conference.

S. Ding , M. Remerova, Rob van der Mei, B. Zwart. 25 November 2013. Fluid approximation of a Call Center model with redials and reconnects. Cornell University Library,arXiv : 1311.6248v1 [Math.PR].

D. Down and S.P. Meyn. (1997). Piecewise linear test functions for stability and instability of queueing networks. Queueing Systems. 27, 205-226.

N. Ebrahimi. (2006). System reliability based on system wear. Stochastic Models, 22(1) : 21-36.

L. C. Evans. (1998). Partial differential equations. Graduate Studies in Mathematics 19. American Mathematical Society, Providence, Rhode Island.

G.I. Falin. (1988). On a multiclass batch arrival retrial queue. Adv. Appl. Probab. 20 : 483-487.

G. Falin. (1990). A survey of retrial queues. Queueing Systems : Theory and Applications, 7(2) : 127-167.

G. I. Falin and J. R. Artalejo. (1998). An infinite source retrial queue. European Journal of Operational Research, 108(2) : 409-424.

N Gans, G Koole, and A Mandelbaum. (2002). Telephone calls centers : a tutorial and literature review. European Jour. Oper. Res.

N. Gharbi and M. Ioualalen. (2006). GSPN analysis of retrial systems with server break- downs and repairs. Applied Mathematics and Computation, 174(2) : 1151-1168.

O. Garnett, A. Mandelbaum, and M. Reiman. (2002). Designing a call center with impatient customers. Manufacturing & Service Operations Management, 4(3) : 208-227.

R. Getoor. (1980). Transience and recurrence of Markov processes. In Azéma, J. and Yor, M. (Eds.), Séminaire de Probabilités XIV, 1978/79, Lecture Notes in Mathematics 784. Springer-Verlag Berlin Heidelberg, pp. 397Ű409.

S.A. Grishechkin. (1992). Multiclass batch arrival retrial queues analyzed as branching processes with immigration. Queueing Syst, 11 : 395-418.

S. Halfin and W. Whitt. (1981). Heavy-traffic limits for queues with many exponential servers. Operations Research, 29(3) : 567-588.

D. Hinrichsen and A. J. (2005). Pritchard. Mathematical systems theory. I. Modelling, state space analysis, stability and robustness. Texts in Applied Mathematics 48. Springer, Berlin Heidelberg.

Ye. Heng-Qing . (2007). A paradox for admission control of multiclass queueing network with differentiated service. J. Appl. Probab., 44(2), 321-331.

R. Ibrahim and W. Whitt. (2009). Real-time delay estimation based on delay history. Manufacturing & Service Operations Management, 11(3) : 397-415.

R. Ibrahim and W. Whitt.(2011). Wait-time predictors for customer service systems with time-varying demand and capacity. Operations Research, 59(5) : 1106-1118.

H. Kaspi and A. Mandelbaum. (1992). Regenerative closed queueing networks. Stochastics Stochastics Rep., 39(4) : 239-258.

V.G. Kulkarni. (1986). Expected waiting times in a multiclass batch arrival retrial queue. J. Appl. Probab, 23 : 144-154.

S. Kumar and P. R. Kumar. (1994). Performance bounds for queueing networks and scheduling policies. IEEE Transactions on Automatic Control 39 : 1600-1611.

P. R. Kumar and S. P. Meyn. (1995). Stability of queueing networks and scheduling policies. IEEE Transactions on Automatic Control 40 : 251-260.

P. R. Kumar and S. P. Meyn. (1996). Duality and linear programs for stability and performance analysis of queuing networks and scheduling policies. IEEE Transactions on Automatic Control 41 : 4-17.

P. Kumar and T. I. Seidman. (1990). Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. IEEE Trans. Autom. Control, 35(3) : 289-298.

C. Langaris, I. Dimitriou. (2010). A queueing system with n-phases of service and $(n-1)-$types of retrial

customers. Eur. J. Oper. Res, 205 : 638-649.

L. Libman and A. Orda. (2002). Optimal retrial and timeout strategies for accessing network resources. IEEE/ACM Transactions on Networking, 10(4) : 551-564.

A. Mandelbaum, W. Massey. and M. Reiman. (1998). Strong approximations for markovian service networks. Queueing Systems, 30(1-2) : 149-201 .

A. Mandelbaum, W. Massey, M. Reiman, and B. Rider. (1999). Time varying multiserver queues with abandonment and retrials. In Proceedings of the 16th International Teletraffic Conference, 4 : 4-7.

A. Mandelbaum, W. Massey, M. Reiman, A. Stolyar, and B. Rider. (2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. Telecommunication Systems, 21(2-4) : 149-171.

S. Meyn. (2008). Control techniques for complex networks. Cambridge University Press, New York.

S. P. Meyn. (1995). Transience of multiclass queueing networks via fluid limit models. Ann. Appl. Probab., 5(4) : 946-957.

S. Meyn and R. Tweedie. (1993). Markov chains and stochastic stability. Springer, London.

E. Moutzoukis, C. Langaris. (1996). Non-preemptive priorities and vacations in a multiclass retrial queueing system. Stoch Models, 12(3) : 455-472.

P. Nain. (1985). Analysis of a two-node Aloha network with infinite capacity buffers. In : T. Hasegawa, H. Takagi, Y. Takahashi, (eds.) Proc. Int. Seminar on Computer Networking and Performance Evaluation. Tokyo, Japan, 18-20.

G. Pang, R. Talreja, and W. Whitt. (2007) Martingale proofs of many-server heavy-traffic limits for markovian queues. Probability Surveys, 4 : 193-267.

A. Papoulis. (1983). Probability Random Variables and Stochastic Processes, 2nd ed., McGraw-Hill.

A. Pukhalskij and A. Rybko. (2000). Nonergodicity of a queueing network under nonstability of its fluid model. Probl. Inf. Transm., 36(1) : 23-41.

A. Puhalskii and A.N. Rybko. (2000). Non-ergodicity of queueing networks under nonstability of their fluid models. Problems of information transmission, 36(1), 26-48.

A. Rybko and A. Stolyar. (1992). Ergodicity of stochastic processes describing the operation of open queueing networks. Probl. Inf. Transm, 28(3) : 199-220.

A.N. Rybko and A.L. Stolyar. (1992). Ergodicity of stochastic processed describing the operations of open

queueing networks. Problemy Peredachi Informatsii, 28, 2-26.

A. N. Rybko and A. L. Stolyar. (1993). On the ergodicity of random processes that describe the functioning of open queueing networks. Problems of Information Transmission 28 : 199-220.

A. N. Rybko and A. L. Stolyar. (1992). On the ergodicity of random processes that describe the functioning of open queueing networks. Problems on Information Transmission, 28, no. 3, 3-26.

T. I. Seidman. (1994). First come, first served can be unstable. IEEE Trans. Autom. Control, 39(10) : 2166-2171.

M. Sharpe. (1988). General theory of Markov processes. Academic Press, Inc., San Diego.

A. Stolyar. (1995). On the stability of multiclass queueing networks : A relaxed sufficient condition via limiting fluid processes. Markov Process. Relat. Fields, 1(4) : 491-512.

D. Sze. (1984). Or practice - a queueing model for telephone operator staffing. Operations Research, 32(2) : 229-249.

W. Szpankowski. (1994). Stability conditions for some multiqueue distributed systems : buffered random access systems. Adv. Appl. Probab, 26 : 498-515.

J. Walrand. (1991). Communication Networks : A First Course. The Aksen Associates Se- ries in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA.

W. Whitt. (2006). Fluid models for multiserver queues with abandonments. Operations Research, 54(1) : 37-54.

H. Q. Ye and H. Chen. (2001). Lyapunov method for the stability of fluid networks. Operations Research Letters, 28(3) : 125-136.

Abstract

In this thesis we establish a sufficient condition for the stability of a queueing network under priority service discipline composed of N-units, N greater than 3 and N^2 classes, and this by using the piecewise linear Lyapunov function. Then we study the stabilization of another model of a queueing network under priority service discipline; A multi-class fluid queuing system with priority composed of N stations, N greater than 3 and 2N classes. In this work the fluid model approach is employed in the study of the stability. Finally we analyze a call center system with two reconnect orbits, two redial (retrial) orbits, "s" servers and two independent Poisson streams of customers, in this work the fluid model is used to derive a first order approximation for the number of customers in the redial and reconnect orbits.

Résumé

Dans cette thèse nous étudions la stabilité de différents systèmes de files d'attente fluides multi-classes. Nous donnons une condition nécessaire de la stabilité d'un système à N stations et N^2classes de clients sous des disciplines de service avec priorité, N supérieur ou égale à 3 et ceci en utilisant la fonction de Lyapunov linéaire par morceau. Ensuite nous étudions la stabilisation d'un autre modèle de réseaux de files d'attente sous des disciplines de service avec priorité; un système de files d'attente fluide multi-classe avec priorité composé de N stations, N supérieur à 3 et 2N classes. Nous nous basons sur le modèle fluide pour l'étude de la stabilité. Et nous terminons par l'analyse d'un système de centre d'appel composé de deux orbites réservée aux clients impatients dite "redial orbit", et deux autres réservés aux clients revenant de l'extérieur appelée "reconnect orbit", s serveurs et deux catégories de clients suivant un processus de Poisson. Dans ce travail, le modèle fluide est utilisé pour calculer une approximation du premier ordre du nombre de clients dans les orbites du système.

ملـــخـــــص

في هذه الأطروحة تمت دراسة استقرار أنظمة مختلفة لصفوف الزبائن في حالة الإنتظار أمام محطات متعددة.في البداية نعطي شرط ضروري لاستقرار أحد الانظمة التى تحتوى على نون محطة و نون مربع فئة من الزبائن, الذين يمرون حسب الاولوية المحددة في الاول و ذلك باستعمال دالة ليبابونوف الخطية. ثم ندرس استقرار نظام من نوع آخر لشبكات ذات صفوف الانتظار زبائنها يمرون حسب الاولوية المحددة في الاول. هذه الشبكة تحتوى على نون محطة, بحيث نون اكبر او يساوي 3 و 2 نون فئة من الزبائن الذين يمرون حسب الاولوية. و في النهاية نقوم بتحليل نظام من نوع مركز اتصال مكون من مدارين للزبائن الذين لا يستطيعون الانتظار و مدارين اخرين للزبائن القادمين من الخارج و عدد معين من الخوادم و نوعين من الزبائن . في هذا العمل, تم استخدام نموذج معين لحساب العدد التقريبي للزبائن في مدارات النظام.