

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITES DE BLIDA 1

FACULTE DES SCIENCES

Département de Mathématiques



Mémoire de fin de cycle en vue d'obtention du diplôme de Master
en Mathématiques

Spécialité : Modélisation Stochastique et Statistique

Thème :

*Analyse de file d'attente markovienne avec
priorités (relative et absolue)*

Présenté par:

✦ M^{elle} COULIBALY Aïché

✦ M^{elle} KADI Nafissa

Devant le jury composé de :

Président :	M ^r O. TAMI	MCB	USDB 1
Examinatrice :	M ^{elle} A. BOUKAFTANE	MAA	USDB 1
Promotrice :	M ^{me} N. OUKID	Prof	USDB 1

Année universitaire : 2021/2022

Remerciement

Nous tenons en premier lieu à remercier Allah Omnipotent, Omniprésent et Omniscient, qui nous a orienté au chemin du savoir et les portes de la science et qui nous a donné la santé, le courage et la volonté pour arriver à bout de ce travail.

Nous remercions du fond du cœur notre promotrice M^{me}.N.Oukid, pour son suivi, sa disponibilité, ses conseils qui nous ont permis de mener à bien ce travail.

Nos vifs remerciements à nos familles pour nous avoir aidées à surmonter tous les obstacles et forger sur dent à travers les difficultés vécues tout au long de cette période de travail.

Nos chaleureux et sincères remerciements vont aussi, à tous les enseignants et l'administration pour leur amitié, leur aide, et leur soutien. Ces années que nous avons passées ensemble, furent très agréables.

Nous exprimons aussi notre profonde gratitude aux membres du jury, de nous avoir fais l'honneur d'examiner, d'évaluer et de juger ce modeste travail.

Dédicace

A ceux qui m'ont tout donné sans rien demander, qui m'ont offert leur soutien durant toute ma vie, qui m'ont aidé sans restriction, qui m'ont appris les bons principes, à qui tous les mots du monde n'arriveront jamais à les décrire.

À l'esprit de mon père qui m'a beaucoup soutenu dans ce que je suis aujourd'hui.

A ma source de courage, à ce que j'ai de plus cher ma mère, que Dieu ait d'elle sa miséricorde.

A mes chères sœurs (Soumia, Khansa, Imen) et mes chers frères (Sofien, Yacine, Anes et sohaybe). A ma grande mère à qui je souhaite une longue vie pleine de bonheur.

A mes amies (Ikram, Anissa, Sara, Wafia, Nour) et ma belle binôme Aiché qui m'a accompagné dans ce travail.

A toute ma famille...

A toutes mes amies...

Dédicace

Avec l'expression de ma reconnaissance, je dédie cet ouvrage

A mon repère, l'école de mon enfance, celui qui m'a arrosé de tendresse et d'espoirs, à toi grand père « Sambou Diakité » puisse le tout puissant t'accorde une longue vie dans la santé.

A mes sources de vie, mes parents recevez ma gratitude et mon affection.

A mes exemples éternels, mes tantes (Ami, Mah, INA) pour leur conseil, leur encouragement et leur soutien tout au long de mes études.

A mes oncles (Allaye, Dramane) et leurs petites familles.

A toutes les personnes de ma grande famille.

A la communauté Malienne de Blida, particulièrement mes sœurs de la Cité 05.

A ma chère binôme pour sa patience.

A toutes mes amies, au nom de l'amitié qui nous réunit, et au nom de nos souvenirs inoubliables.

ملخص:

في هذا العمل ، نحن مهتمون بنمذجة انقطاع الخدمة بسبب زبون يتمتع بالأولوية في أنظمة الانتظار. نتيجة لذلك ، قمنا أولاً بمراجعة أنظمة الانتظار الكلاسيكية الماركوفية ونصف الماركوفية ، ثانياً قمنا بدراسة تأثير بعض عوامل إنقطاع الخدمة بصفة عامة (على سبيل المثال : وصول سلبي ، خادم خارج الخدمة ، وخادم في عطلة ...). أخيراً، أجرينا تحليلاً عشوائياً لنظام قائمة الانتظار مع الأولوية (المطلقة والنسبية) واختتمناها بمقارنة بين الأولويتين.

الكلمات المفتاحية:

خط الإنتظار, قياس الكفاءة, أولوية مطلقة, أولوية نسبية, معادلة الموازنة, دالة التوليد.

Résumé

Dans ce travail, nous nous sommes intéressés à la modélisation de l'interruption de service due à une arrivée jouissante d'une priorité dans les systèmes d'attente. De ce fait, nous avons passé en revue premièrement les systèmes d'attente classiques markoviens et semi-markoviens, puis en second lieu étudié globalement l'influence de certains facteurs d'interruption (des arrivées négatives, des pannes de serveur, des vacances simples de serveur...). Finalement, nous avons réalisé une analyse stochastique du système de file d'attente M/M/1 avec priorité (absolue et relative) puis procédé à une comparaison entre les deux.

Mots clés :

Files d'attente, mesures de performance, priorité relative, priorité absolue, équations de balance, fonctions génératrices.

Abstract

In this work, we are interested in the modeling of the interruption of service due to the arrival enjoying a priority in the waiting systems. As a result, first we reviewed the classic Markovian and semi-Markovian waiting systems, then, secondly, we studied globally the influence of certain interruption factors (negative arrivals, server failures, simple vacancies server...). Finally, we performed a stochastic analysis of the M/M/1 queuing system with priority (absolute and relative) and then proceeded to a comparison between the two.

Key word:

Queues, performance measurement, preemptive priority, non-preemptive priority, balance equation, generating function.

Table des matières

TABLE DE FIGURE	
LISTE DES TABLEAUX	
INTRODUCTION GENERALE	1
CHAPITRE1 : SYSTEMES DE FILE D'ATTENTE
Introduction	4
1.1 Définitions	4
1.2 Description d'un système de file d'attente simple	4
1.3 Classification des systèmes d'attente	5
1.3.1 Processus des arrivées des clients	5
1.3.2 Processus de service	5
1.3.3 Structure et discipline de la file	6
1.4 La notation du Kendall	7
1.5 Définitions	8
1.5.1 Processus stochastique	8
1.5.2 Chaîne de Markov :	8
1.6 Analyse mathématique des systèmes de files d'attente	9
1.6.1 Les caractéristiques d'exploitation ou mesures de performances d'un système	10
1.6.2 Les formules de Little	11
1.7 Processus de naissance et de mort	11
1.8 Caractéristiques des systèmes de files d'attente markoviennes	12
1.8.1 Système d'attente M /M/1	13
1.8.2 Système M/M/S	15
1.8.3 Système M/M/ ∞	18
1.8.4 Système d'attente M/M/1/K	20
1.9 Description d'un système semi markovien	23
1.9.1 Système M/G/1	23
1.9.2 Système G/M/1	27
Conclusion	29

CHAPITRE2 :INTERRUPTION DE SERVICE DANS LES SYSTEMES D'ATTENTE

Introduction :	30
2.1 Système de file d'attente avec vacance de serveur :	31
2.1.1 Les caractéristiques des politiques de vacances	32
2.1.2 Notation	33
2.1.3 Modèle $M/M/1/\infty/\infty/Eu$	34
2.2 Système d'attente avec panne du serveur	38
2.2.1 Pannes passives	38
2.2.2 Pannes actives	39
2.3 Systèmes de files d'attente avec arrivées négatives	40
2.3.1 Le processus d'élimination des clients positifs	41
2.3.2 Notation	41
2.3.3 Les différents types de clients négatifs	42
2.4 Système d'attente avec priorité	43
2.4.1 priorité relative	43
2.4.2 Priorité absolue	44
Conclusion	45

CHAPITRE3 :L'ANALYSE DU SYSTEME D'ATTENTE M/M/1 AVEC PRIORITE RELATIVE ET ABSOLUE.....

Introduction	46
3.1 Description du Système d'attente M/M/1 avec priorité	46
3.1.1 Système d'attente M/M/1 avec Priorité relative	48
3.1.2 Système d'attente M/M/1 avec Priorité absolue	62
3.2 Les illustrations numériques :	70
3.2.1 L'effet de la variation du taux λ_1 sur le temps moyen d'attente des clients dans la file pour les deux systèmes (M/M1 avec priorité relative et M/M/1 avec priorité absolue)	70
3.2.2 L'effet de la variation du taux λ_2 sur le temps moyen d'attente des clients dans la file pour les deux systèmes (M/M1 avec priorité relative et M/M/1 avec priorité absolue)	71
3.2.3 Interprétations	73
CONCLUSION GENERALE ET PERSPECTIVES	74
Bibliographie	75

Table de figure

Les figures	Sa notation
Figure 1	Mécanisme d'un système de file d'attente simple.....5
Figure 2	Graphe de transition du processus de naissance et de mort.12
Figure 3	Le système markovien de file d'attente.12
Figure 4	Mécanisme de fonctionnement du système M/M/1. 13
Figure 5	Graphe de transition du système d'attente M/M/113
Figure 6	Représentation schématique d'une file M/M/∞.16
Figure 7	Graphe de transition d'un système d'attente M/M/S.16
Figure 8	Graphe de transition d'un système d'attente M/M/∞19
Figure 9	Le graphe de transition d'un système M/M/1/K.21
Figure 10	les systèmes semi markoviens.23
Figure 11	Schéma représentatif d'un système d'attente avec vacance du serveur.31
Figure 12	Graphe de transition d'un système d'attente avec vacance.35
Figure 13	Schéma représentatif d'une panne passive.39
Figure 14	Schéma représentatif d'une panne active.39
Figure 15	Variation de W_{q_1} (absolu et relative) en fonction de λ_170
Figure 16	Variation de W_{q_2} (absolu et relative) en fonction de λ_171
Figure 17	Variation de W_{q_1} (absolu et relative) en fonction de λ_272
Figure 18	Variation de W_{q_2} (absolu et relative) en fonction de λ_272

Liste des tableaux

Les tableaux	Son description
Tableau 1	les différents codes d'un processus d'arrivée et de service..... 6
Tableau 2	Les notations possibles de la discipline. 7
Tableau 3	Les mesures de performance d'un système M/M/1. 15
Tableau 4	Les mesures de performance d'un système M/M/S. 18
Tableau 5	Les mesures de performance d'un système M/M/∞..... 20
Tableau 6	Les mesures de performance d'un système M/M/1/K. 23
Tableau 7	Les mesures de performance d'un système M/G/1. 26
Tableau 8	Les mesures de performance d'un système G/M/1. 28
Tableau 9	L'effet de λ_1 sur W_{q1} (absolu et relative). 70
Tableau 10	L'effet de λ_1 sur W_{q2} (absolu et relative). 71
Tableau 11	L'effet de λ_2 sur W_{q1} (absolu et relative). 72
Tableau 12	L'effet de λ_2 sur W_{q2} (absolu et relative). 72

INTRODUCTION GENERALE

Dans le quotidien, nous passons une partie de notre temps dans la file d'attente, pourtant il est évident que faire la queue est l'un des irritants de notre vie, nous l'acceptons uniquement que si nous y sommes contraints et forcés.

Ainsi la manifestation des phénomènes d'attente de la vie courante peut se faire sous diverses formes comme par exemple : la vente de billets auprès d'un guichet, la gestion des trafics aériens, l'arrivée des voitures vers une station de service, l'exécution des tâches dans un centre de calcul, etc.

Ces phénomènes d'attente peuvent être modélisés et étudiés par la théorie des files d'attente. En effet cette théorie est une vaste branche des sciences. Très rapide en développement, et appartenant à la théorie de probabilités et à la modélisation stochastique, l'origine de la théorie des files d'attente remonte il y a bien longtemps avec les travaux précurseurs d'Anger Krarup Erlang (1909) sur l'analyse du modèle pour la communication téléphonique à l'aide du processus de Poisson.

Plus tard ces résultats furent encore étendus dans différentes directions par les travaux des mathématiciens célèbres tels que Pollaczek aux Etats Unis, Khintchine en Russie, Kendall en Cambridge, Kleinrock et bien d'autres.

Les chercheurs travaillant dans cette branche d'activité ont élaborés plusieurs nouvelles méthodes qui ont été ensuite appliquées avec succès dans d'autres domaines, particulièrement dans le secteur de la fabrication. On a également constaté une résurgence des applications pratiques de la théorie des files d'attente dans des secteurs plus traditionnels de la recherche opérationnelle, un mouvement mené par Peter Kolesar et Richard Larson. Grâce à tous ces développements, la théorie des files d'attente connaît aujourd'hui une croissance rapide dans divers champs, notamment une analyse théorique des modèles de file d'attente et des réseaux d'une structure plutôt complexe en utilisant des modèles mathématiques assez sophistiqués et divers types de processus stochastiques. Elle englobe également des domaines très étendus d'applications modernes : l'informatique, les télécommunications en général, l'ingénierie des trafics, la fiabilité des systèmes complexes

Les diverses situations dont la modélisation et l'analyse s'effectuent par la théorie des files d'attente suivent en général le même schéma descriptif suivant : les clients arrivent de l'extérieur (source) à des instants aléatoires, ils attendent que l'un des serveurs soit libre pour

pouvoir être servi puis quittent le système. La durée du service auprès de chaque serveur est elle-même aléatoire.

Pour spécifier un système d'attente, il est primordial de définir les éléments suivants : le processus des inter-arrivées et des durées de service des clients, la capacité de la file, ainsi que la discipline de service, etc.

La plupart des systèmes d'attente supposent l'arrivée d'une classe unique de clients au sein du système, mais la diversité des systèmes nous permette donc de prendre en considération les systèmes avec plusieurs types de clients ou certains jouissent d'une priorité de service, un tel système est connu sous le nom de système d'attente avec priorité. Ainsi la nature de cette priorité peut être absolue ou relative.

C'est dans le but d'étudier et de comprendre l'influence du phénomène de priorité sur les performances du système que des multitudes de publications ont été faites par différents chercheurs. Ainsi le premier article publié concernant les systèmes d'attente avec priorité fut l'article de Cobham [2] pour la priorité relative et celui de White et Christie [6] pour la priorité absolue. Puis s'en est suivit les articles de Asha Seth Kapadia [], Stephan [3], Douglas R Miller [5], Jaiswal [9], Holley [8], Dressin et Reich [10],

L'objectif de notre travail, est d'élargir le champs d'application des systèmes d'attente avec priorité. En particulier, nous nous intéressons donc à une étude analytique détaillée du modèle d'attente M/M/1 avec priorité (relative et absolue), puis nous accompagnons cette étude par des illustrations numériques qui nous permettrons donc d'observer l'influence des différents taux d'arrivées des clients sur certaines caractéristiques du système.

Le reste de ce document est structuré de la manière suivante:

- ✓ Le premier chapitre est basé sur l'étude des systèmes d'attente markoviens (M/M/1, M/M/1/K, M/M/S, M/M/∞) et semi-markoviens (M/G/1, G/M/1).
- ✓ Le deuxième chapitre est consacré à une étude analytique de quelques facteurs d'interruption de service (vacance du serveur, panne du serveur, les arrivées négatives et les priorités) portant sur les systèmes d'attente tout en mettant en relief les différentes politiques (d'arrivée et de service), ainsi que leurs effets sur le service d'un client.

- ✓ Et enfin, dans le troisième chapitre on s'intéresse aux systèmes d'attente markoviens avec priorité. Pour cela, nous analysons le modèle M/M/1 à deux classes avec priorité (relative et absolue) et ceci en passant par la détermination des fonctions génératrices à l'aide des équations de Kolmogorov obtenues grâce aux différents graphes de transition d'état de ce système; puis on déterminera également leurs mesures de performances. Par la suite nous procédons à des illustrations numériques dans le but d'observer l'influence de la variation des différents taux arrivées.

1. Systèmes de file d'attente

Introduction :

La théorie des files d'attente, en anglais « Queueing theory », est la branche des mathématiques qui a pour sujet d'étude la modélisation d'un système, le calcul de ses performances, la détermination de ses caractéristiques (le nombre des clients en attente, le temps d'attente, le temps de séjour ...) et l'évaluation du rendement tout en améliorant le fonctionnement du système afin d'éclairer les praticiens dans leur prises de décisions.

De 1953 à nos jours grâce à l'effort conjugué de certains mathématiciens (Kendall, Kolmogorov, Jackson, Khintchine, Saaty, Palm...), les files d'attentes sont de plus en plus importantes grâce à leur utilisation dans plusieurs domaines et sous divers aspects on peut citer entre autre la gestion d'un fichier informatique, la gestion d'un stock, l'attente devant un guichet de banque etc.

Dans ce chapitre, après quelques notions préliminaires utiles pour notre travail, nous présentons une analyse détaillée des systèmes classiques (markoviens et semi-markoviens).

1.1 Définitions

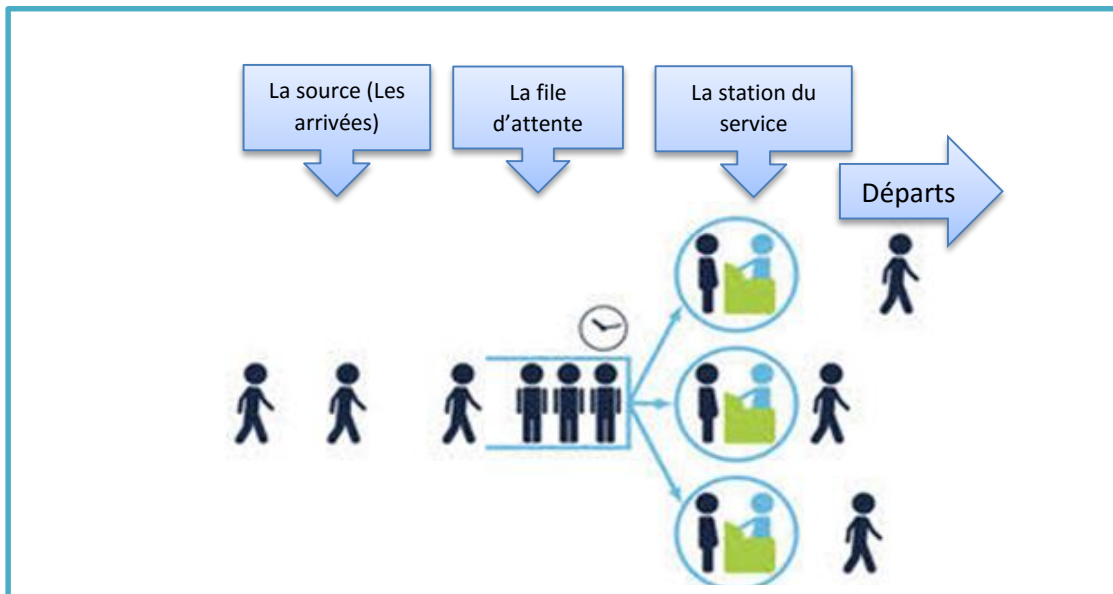
File d'attente : c'est l'ensemble formé par un nombre maximum permis de clients en attente d'être servis (fini ou infini) lorsque le serveur est occupé pendant un certain temps par le service d'un client.

L'état d'un système : Nous appelons état d'un système à l'instant t le nombre de clients présents dans le système à cet instant (un client est présent dans le système s'il est en file d'attente ou en cours de service).

Système d'attente : c'est l'ensemble formé par les clients qui sont dans la file d'attente et ceux qui sont en service.

1.2 Description d'un système de file d'attente simple

Une file d'attente est un système constitué d'un espace d'attente qui contient une ou plusieurs places, et un espace de service composé d'un ou plusieurs serveurs. Les clients arrivent de l'extérieur (source) à des instants aléatoires, ils attendent que l'un des serveurs soit libre pour pouvoir être servi puis quittent le système.



✚ **Figure 1** : Mécanisme d'un système de file d'attente simple

Un système d'attente comporte plusieurs caractéristiques. Typiquement, une file est composée de clients se succédant et demandant un service (voir Fig. 1). Les clients peuvent être des individus, des appels téléphoniques, des signaux électriques, des véhicules, des accidents, des turbulences atmosphériques... et le serveur peut être humain, un central téléphonique, un serveur informatique, une compagnie....

1.3 Classification des systèmes d'attente

La spécification d'une file d'attente simple nécessite donc la caractérisation des éléments suivants [15]:

1.3.1 Processus des arrivées des clients

On s'intéresse aux instants d'arrivées des clients dans la file. Ils sont en général aléatoires. Certaines hypothèses sont faites sur leurs lois. Tout d'abord, il n'arrive qu'un client à la fois. La deuxième hypothèse est l'homogénéité dans le temps. Cela se traduit par le fait que les temps d'inter-arrivées des clients sont des variables aléatoires de même loi. Ils sont également supposés indépendants. Enfin, la loi des temps d'inter-arrivées est supposée connue.

1.3.2 Processus de service

Les durées de services sont des variables positives indépendantes et de même loi. Elles sont mesurées par les quantités séparant le début et la fin du service des clients.

Les codes les plus courants dans la description du processus des arrivées et celui de service des clients sont résumés dans le tableau suivant :

Symbol nom	La description
M (Markov)	Processus de Poisson les temps inter-arrivées et / ou les durées de service sont exponentiellement distribués.
G (général)	Les Inter arrivées des clients et / ou les durées de service ont une distribution générale. Les inter arrivées peuvent être dépendantes.
Ek (Loi d Erlang)	Ce symbole désigne un processus où les intervalles de temps entre deux arrivées successives et / ou les durées de service sont des variables aléatoires indépendantes et identiquement distribués suivant une loi d'Erlang d'ordre k.
GI (général indépendant)	les Inter arrivées des clients et / ou les durées de service ont une distribution générale (les inters arrivés sont indépendantes et identiquement distribués).
D (Répartition déterministe)	les temps inter arrivées des clients et / ou les temps de service sont constants et toujours les mêmes.

■ **Tableau 1 :** les différents code un processus d'arrivée et de service

1.3.3 Structure et discipline de la file :

- **Nombre de serveurs:**
Une station de service peut contenir un ou plusieurs serveurs montés en parallèle ;
- **Capacité de la file :**
Elle est le nombre maximum de places disponibles pour accueillir les clients en attente de service. Ce nombre peut être fini ou infini ;
- **La source des clients :**
Ou encore appelée population est généralement considérée comme infinie car les modèles réels à capacité assez grande qu'elle modélise sont considérés comme infinie, mais cependant elle peut être également finie ;
- **La discipline de service :**
La discipline de service indique l'ordre dans lequel est traité le client. Le tableau suivant résume quelques règles courantes qui sont adoptées :

Symbole	Nom	La description
FIFO/FCFS	first in first out/first come first served	Premier entré, premier sorti/premier arrivé, premier servi
LIFO/LCFS	Last in first out/last come first served	Dernier entré, premier sorti/dernier arrivé, premier servi
SIRO	Served In Random Order	Les clients sont servis dans un ordre aléatoire sans égard à l'ordre d'arrivée.
QP	Priority service	Les clients sont servis selon leur priorité.
PS	Processor Sharing	Les clients sont servis dans l'ordre déterminé sans tenir compte de l'ordre d'arrivée.

■ **Tableau 2** : Les notations possibles de la discipline.

1.4 La notation de Kendall :

La notation de Kendall est le système standard utilisé pour décrire et classer un système de file d'attente, elle est de cette forme [15] :

$$A/B/C/D/E/F$$

Telle que :

- A : Un code décrivant le processus d'arrivée ;
- B : décrit la distribution des temps de service d'un client ;
- C : le nombre de serveur ;
- D : La capacité du système (c'est le nombre maximum de clients autorisé dans le système, c'est à dire Lorsque le nombre de client atteint ce maximum, d'autres arrivées sont refusées) ;
- E : la capacité de la source ;
- F : La discipline de service ;

Remarque :

Quand la capacité du système et celle de la source sont infinies, la discipline de la file est FIFO on se contente des trois premiers symboles.

Avant d'entamer la partie d'analyse mathématique des systèmes de file d'attente, nous allons d'abord définir le processus stochastique et la chaîne de Markov qui sont les deux notions fondamentales en théorie des files d'attente [15].

1.5 Définitions

1.5.1 Processus stochastique

L'étude des processus aléatoires constitue l'un des objectifs les plus profonds dans la théorie des probabilités car les problèmes mathématiques intéressants et souvent très difficiles sont solutionnés par elles. Un processus stochastique $\{X(t)\}_{t \in T}$ est une famille de variables aléatoires dont l'ensemble des temps T peut être discret ou continu. Soit $X(t)$ une variable aléatoire qui définit l'état du processus à chaque instant $t \in T$ donné et E l'espace d'états (ensemble des valeurs que peut prendre le processus à chaque instant) [1].

- Un processus $\{X(t)\}_{t \in T}$ tel que $X(0) = 0$ est à accroissements indépendants si, pour toute suite finie $0 < t_1 < t_2 < \dots < t_n$, les variables aléatoires $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendants ;
- Un processus à accroissements indépendants est à accroissements stationnaires si la loi de l'accroissement $X_{t+s} - X_t$ ne dépend pas de t , pour tout $t \geq 0$;
- Un processus est dit stationnaire (ou homogène dans le temps), si pour tout s et pour tout t ; l'accroissement $X_{t+s} - X_s$ a même loi que X_t ;

Ainsi les processus stochastiques peuvent être différenciés selon T et E comme suit :

- Processus à temps discret et à espace d'états discret ;
- Processus à temps discret et à espace d'états continu ;
- Processus à temps continu et à espace d'états discret ;
- Processus à temps continu et à espace d'états continu ;

1.5.2 Chaîne de Markov

Un processus stochastique $\{X(t)\}_{t \in T}$ est dit chaîne de Markov si et seulement si :

$$P[X_{t+1} = j / X_0 = a_0, X_1 = a_1, \dots, X_t = i] = P[X_{t+1} = j / X_t = i].$$

Le futur état ne dépend que de l'état actuel.

$$\text{On pose : } P_{i,j} = P[X_{t+1} = j / X_t = i].$$

Les $P_{i,j}$ sont appelées probabilités de transition de la chaîne de Markov. La matrice carrée dont les éléments sont les $P_{i,j}$, ($i, j \in E$) est dite matrice de transition de la chaîne de Markov et elle vérifie les propriétés suivantes :

- 1) $0 \leq P_{i,j} \leq 1$, $\forall i, j \in E$
- 2) $\sum_{j \in E} P_{i,j} = 1$ $\forall i \in E$

Alors si ces deux conditions sont vérifiées dans une matrice, elle est donc dite matrice stochastique.

1.6 Analyse mathématique des systèmes de files d'attente :

Le plus souvent, on étudie un système de file d'attente en introduisant un processus stochastique approprié décrivant l'évolution du système et ceci de façon temporaire. Afin d'obtenir les mesures de performances du système, on procédera d'abord par le calcul des probabilités d'état définissant le régime transitoire ou en passant par la distribution stationnaire. [15]

1. Les probabilités d'état $\pi_n(t) = \pi([X_t = n])$ qui définissent le régime transitoire du processus $\{(X_t), t \geq 0\}$. Il est évident que les fonctions $\pi_n(t)$ dépendent de l'état initial ou de la distribution initiale du processus.

2. Le régime stationnaire du processus, défini par

$$\pi_n = \lim_{n \rightarrow +\infty} \pi_n(t) = \lim_{n \rightarrow +\infty} \pi(X_t = n) \quad n=1, 2, 3, \dots$$

Où π_n est appelée distribution stationnaire du processus $\{X_t, t \geq 0\}$.

Par la suite on se contentera de déterminer le régime stationnaire du système car le calcul explicite du régime transitoire semble difficile voire impossible pour les modèles en général mis à part pour certains modèles particulièrement faciles à traiter.

L'intérêt principal du calcul de la distribution stationnaire se porte sur le nombre de client $X(t)$, se trouvant dans le système à l'instant t ($t \geq 0$) car ce dernier permet de déterminer le nombre de client servit et le nombre de client en attente de service, il permet également de faire une déduction de toutes les valeurs moyennes des autres paramètres.

1.6.1 Les caractéristiques d'exploitation ou mesures de performances d'un système :

Chaque modèle de file d'attente est défini par ses performances, ainsi l'objectif principal de l'étude d'une file d'attente ou d'un réseau de file d'attente est le calcul ou l'estimation des performances d'un système dans des conditions de fonctionnement données. Les performances les plus utilisées sont :

- ✓ Le nombre moyen de clients dans le système ;
- ✓ Le nombre moyen de clients dans la file d'attente ;
- ✓ Le temps moyen de séjour d'un client dans le système (attente + service) ;
- ✓ Le temps moyen d'attente d'un client dans la file ;

Notations [15]

λ : Taux des arrivées des clients ;

λ_e : Taux d'effectif d'entrée dans le système ;

μ : Taux de service ;

$\frac{1}{\mu}$: Temps moyen de service ;

L : Nombre moyen de clients dans le système ;

L_q : Nombre moyen de clients dans la file d'attente ;

W_q : Temps moyen d'attente d'un client dans la file ;

W : temps moyen de séjour d'un client dans le système. ;

Remarque :

Le calcul des mesures de performance se fait le plus souvent uniquement pour le régime stationnaire.

L' **intensité de trafic** ρ d'un système est une mesure importante qui mesure le degré de saturation de ce système et elle est obtenue par le rapport de la demande (mesurée grâce au taux d'arrivée λ) et la capacité de service (par le taux de service μ).

$$\rho = \frac{\lambda_e}{\mu}$$

1.6.2 Les formules de Little :

La relation liant tous ces paramètres s'appelle **formules de Little**, cette relation n'est acceptable que si $\rho < 1$, et elle est définie par :

$$\begin{aligned} L &= \lambda_e W ; & W &= W_q + \frac{1}{\mu} ; \\ L_q &= \lambda_e W_q ; & L &= L_q + \frac{\lambda_e}{\mu} \end{aligned}$$

Remarque : Dans le cas où la capacité de la file est infini, on a $\lambda_e = \lambda$.

1.7 Processus de naissance et de mort [15] :

Dans un système d'attente, l'évolution temporelle du nombre de clients X_t dans le système à un instant t est décrite par les processus de naissance et de mort qui sont des processus stochastiques à temps continu et à espace d'états discret $n=0, 1, 2, 3, \dots$. Ils sont également sans mémoire et homogènes dans le temps.

$$P(X(t+s)=j/X(s)=i) = P(X(t)=j/X(0)=i) = P_{i,j}(t) \quad (\text{ne dépend pas de } s)$$

$\{X(t), t \geq 0\}$ est un processus de naissance et de mort si :

- $P_{i,i+1}(\Delta t) = \lambda_i \Delta t + o(\Delta t) \quad \forall i \geq 0 ; \quad \lambda_i > 0 ; o(\Delta t)$ infiniment petit.
- $P_{i,i-1}(\Delta t) = \mu_i \Delta t + o(\Delta t) \quad \forall i \geq 1 ; \quad \mu_i > 0 \text{ et } \mu_0 = 0$
- $P_{i,i}(\Delta t) = 1 - (\lambda_i + \mu_i) \Delta t + o(\Delta t) \quad \forall i \geq 1 ;$
- $P_{i,j}(\Delta t) = o(\Delta t)$ pour $|i-j| \geq 2$ (à partir d'un état donnée n)

Ce qui signifie qu'à partir d'un état donné n le passage n'est possible que vers l'un ou l'autre des états voisins $(n-1)$ et $(n+1)$.

Avec :

$$\begin{cases} \lambda_i = \text{taux de naissance (croissance)} \\ \mu_i = \text{taux de mort (décroissance)} \\ P_{i,j} = \text{la probabilité de passer de l'état } i \text{ à l'instant } (t) \text{ à l'état } j \text{ à l'instant } (t + \Delta t) \end{cases}$$

Le graphe de transition de transition du processus de naissance et de mort est :

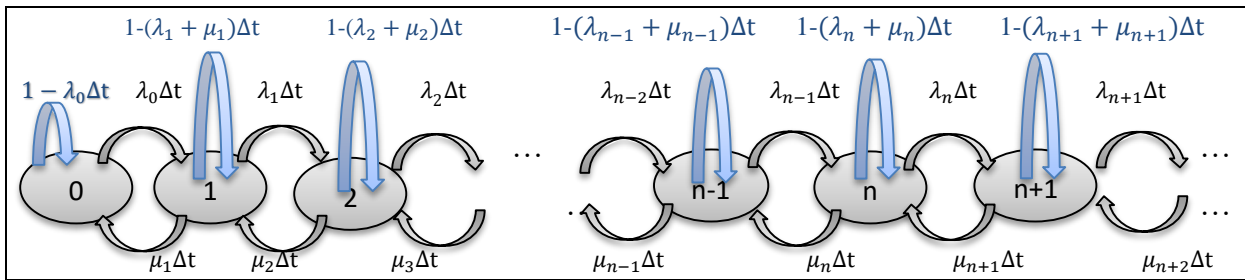


Figure 2 : Graphe de transition du processus de naissance et de mort

1.8 Caractéristiques des systèmes de files d'attente markoviens :

markoviens :

La plupart du temps, l'arrivée des clients à une file simple est supposée décrite par un processus de renouvellement. Le processus d'arrivée le plus simple et le plus couramment employé est le processus de Poisson, d'où la cause de la facilité des modèles markoviens car ces modèles caractérisent les systèmes dans lesquels les inter arrivées et la durée de service sont des variables aléatoires indépendantes exponentiellement distribuées.

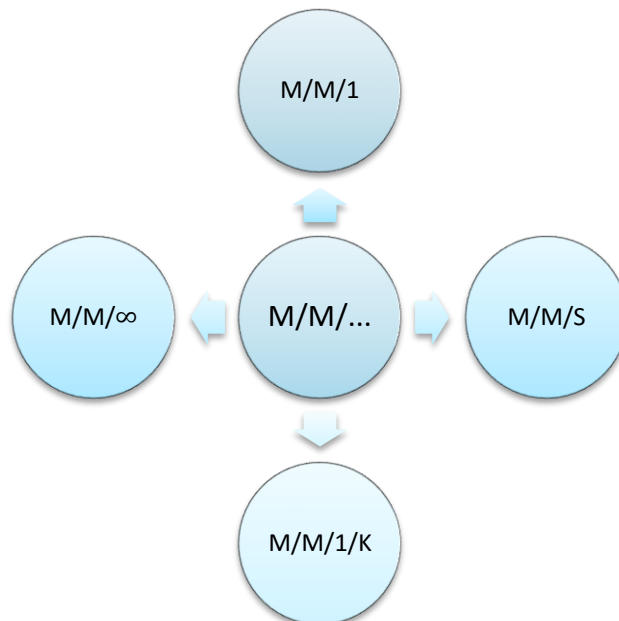


Figure 3 : Le système markovien de file d'attente

1.8.1 Système d'attente M /M/1 :

Le système de file d'attente M/M/1 est le système le plus élémentaire de la théorie des files d'attente. Le flux des arrivées est poissonnien de paramètre λ , la durée de service est exponentielle de paramètre μ , et la discipline d'attente est FIFO, avec une capacité du système et de source infinie. Dans ce système si le serveur est libre le client est forcément pris en charge sinon il rejoint la file d'attente.

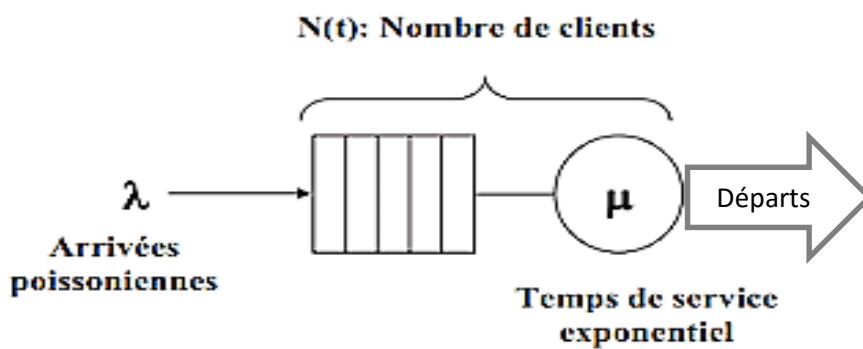


Figure 4 : Mécanisme de fonctionnement du système M/M/1.

Le processus $(N(t); t \geq 0)$ est un processus de naissance et de mort avec un taux de naissance $\lambda_i = \lambda$, pour tout $i \geq 0$ et le taux de mortalité $\mu_i = \mu$ pour tout $i \geq 1$.

$\lambda_i = \lambda$	$i \geq 0$
$\mu_i = \mu$	$i \geq 1$

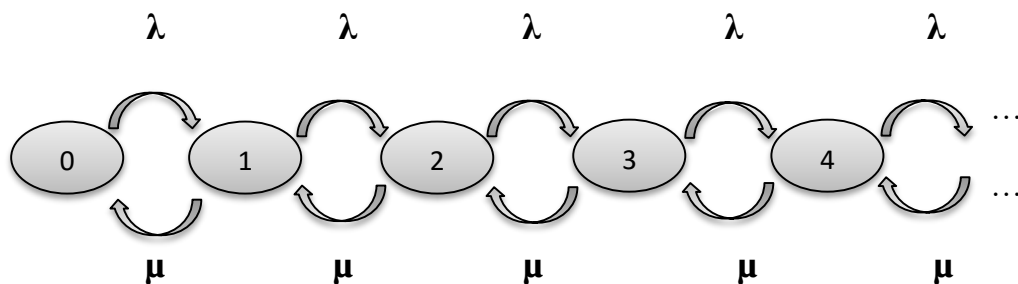


Figure 5 : graphe de transition d'un système d'attente M/M/1.

Distribution stationnaire du système M/M/1

Dans ce modèle pour trouver la probabilité stationnaire π_n il suffit d'appliquer la méthode de résolution des équations de Kolmogorov pour un processus de naissance et de mort.

Utilisons cette formule :

$$\pi_n = \pi_0 \prod_{i=1}^n \frac{\lambda_i - 1}{\mu_i}$$

Nous avons

$$\begin{cases} \lambda_n = \lambda \\ \mu_n = \mu \end{cases} \quad \forall n > 0$$

On pose $\rho = \frac{\lambda}{\mu}$ (qui s'appelle intensité de trafic), D'où le régime stationnaire s'établit si: $\frac{\lambda}{\mu} < 1 \Rightarrow \frac{1}{\mu} < \frac{1}{\lambda}$ qui veut dire que la durée moyenne des services est inférieure à l'intervalle moyen entre deux arrivées. Par l'application de la formule précédente on trouve :

$$\pi_n = \pi_0 \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} = \pi_0 \prod_{i=1}^n \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^n$$

On a $\pi_0 = (1 - \rho)$ si on remplace on obtient :

$$\pi_n = \rho^n (1 - \rho)$$

$\pi = \{\pi_n\}_{n \geq 0}$ est appelé distribution stationnaire, elle suit une loi géométrique.

Mesures de performance d'un système M/M/1

Nombre moyen de clients dans le système :

$$L = E(X) = \sum_{n=0}^{\infty} n \pi_n = (1 - \rho) \sum_{n=0}^{\infty} n \rho^n = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda}$$

Nombre moyen de clients dans la file d'attente :

$$L_q = E(X_q) = \sum_{n=1}^{\infty} (n-1) \pi_n = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Temps moyen de séjour d'un client dans le système :

$$W = \frac{L}{\lambda} = \frac{\lambda}{\lambda(\mu-\lambda)}$$

Temps moyen d'attente d'un client :

$$W_q = w - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$$

L	L_q	W	W_q
$\frac{\lambda}{\mu - \lambda}$	$\frac{\lambda^2}{\mu(\mu - \lambda)}$	$\frac{1}{(\mu - \lambda)}$	$\frac{\lambda}{\mu(\mu - \lambda)}$

■ **Tableau 3** : Les mesures de performances d'un système M/M/1.

1.8.2 Système M/M/S :

Ce système est identique au système M/M/1 excepté qu'il comporte **S serveurs** en parallèles identiques et indépendants les uns des d'autres. On conserve donc les mêmes hypothèses du système M/M/1 (processus d'arrivée des clients poissonien de taux λ , le temps de service exponentiel de taux μ (pour chaque serveur), la discipline d'attente est FIFO, le système et la source sont de capacité infinie). L'accès au serveur s'effectue comme suit :

Si l'un des serveurs est libre, le client qui arrive se dirige immédiatement vers ce serveur, sinon il prend place dans la file d'attente commune à tous les serveurs et ainsi lorsqu'un serveur se libère, le client en tête de la file l'occupe. Pour mieux comprendre le processus de fonctionnement du système M/M/S, nous avons donc effectué une représentation graphique en prenant pour exemple S=8.

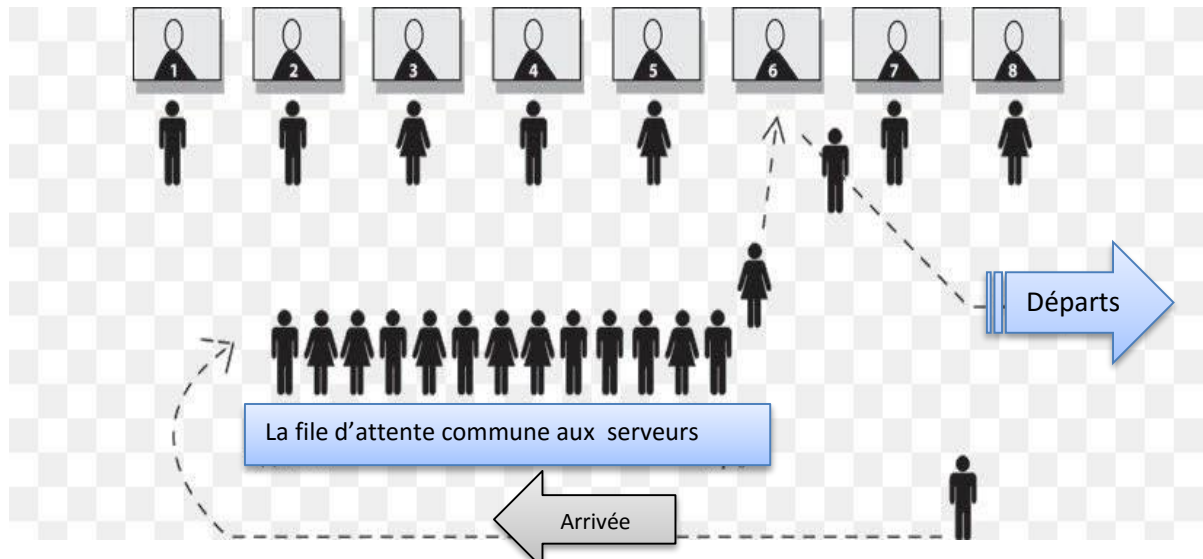


Figure 6 : Représentation schématique d'une file M/M/8.

Dans ce cas aussi, le processus modélisant le nombre de clients dans le système est un processus de naissance et de mort avec :

$$\lambda_n = \lambda \quad n \geq 0$$

$$\mu_n = \begin{cases} n\mu & \text{pour } n = 1, 2, \dots, s-1 \\ s\mu & \text{pour } n \geq s \end{cases}$$

Voici ci-dessous le graphe correspondant à ce système :

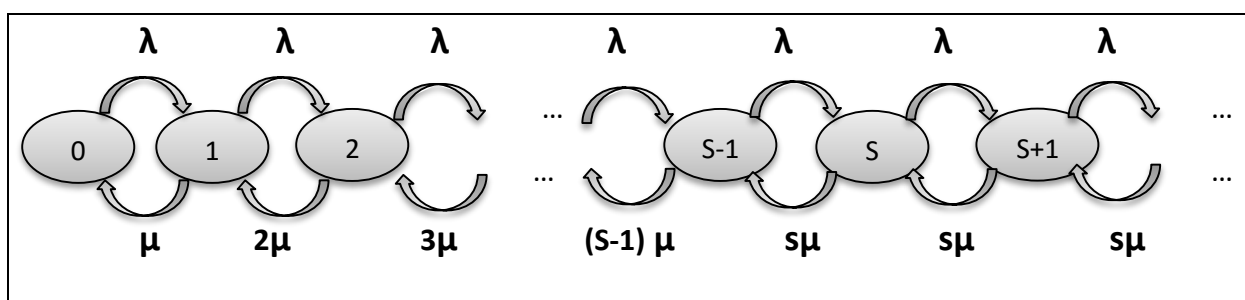


Figure 7 : Graphe de transition d'un système d'attente M/M/S

Distribution stationnaire du système M/M/S :

Pour calculer π_n , le système doit être stable ($\lambda < s\mu$ donc $\rho < 1$; avec $\rho = \frac{\lambda}{s\mu}$)

D'après le graphe de transition et à l'aide de la procédure des équations de Kolmogorov on obtient les équations suivantes :

$$\left\{ \begin{array}{l} \lambda \pi_0 = \mu \pi_1 \\ (\lambda + n\mu) \pi_n = \lambda \pi_{n-1} + (n+1)\mu \pi_{n+1} \quad 1 \leq n < s \\ (\lambda + s\mu) \pi_n = \lambda \pi_{n-1} + s\mu \pi_{n+1} \quad n \geq s \\ \sum_{n=0}^{\infty} \pi_n = 1 \end{array} \right.$$

Où :

π_n représente la probabilité qu'il y ait n clients dans le système.

La résolution du système ci-dessus permet donc d'obtenir la distribution stationnaire suivante :

$$\pi_n = \begin{cases} \pi_0 \frac{\rho^n}{n!} & n = 1, 2, 3, \dots, S-1 \\ \pi_0 \frac{\rho^n}{s^{n-s} s!} & \text{Pour } n \geq s \end{cases}$$

Avec : $\pi_0 = \left[\sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \frac{\rho^s}{s!} \frac{1}{1-\frac{\rho}{s}} \right]^{-1}$.

Mesures de performance d'un système M/M/S :

Nombre moyen de clients dans la file :

$$Lq = \sum_{n=s+1}^{\infty} (n - s) \pi_n = \sum_{n=s+1}^{\infty} (n - s) \frac{\lambda^n}{s^{n-s} s! \mu^n} = \frac{\lambda^s \rho \pi_0}{\mu^s s! (1-\rho)^2}$$

Nombre moyen de clients dans le système :

D'après la formule de Little on a :

$$L = Lq + \frac{\lambda}{\mu} = \frac{\lambda^s \rho}{\mu^s s! (1-\rho)^2} + \frac{\lambda}{\mu} = s\rho + \frac{\rho \pi_0}{(1-\rho)^2}$$

Temps moyen de séjour d'un client dans le système :

D'après la formule de Little on a :

$$W = \frac{L}{\lambda} = \frac{1}{\mu} + \frac{\pi_0}{s\mu(1-\rho)^2}$$

Temps moyen d'attente d'un client :

D'après la formule de Little on a :

$$Wq = W - \frac{1}{\mu} = \frac{\left(\frac{\lambda}{\mu}\right)^s \mu \pi_0}{(s-1)!(s\mu-\lambda)^2}$$

L	Lq	W	Wq
$s\rho + \frac{\rho \pi_0}{(1-\rho)^2}$	$\frac{\lambda^s \rho \pi_0}{\mu^s s! (1-\rho)^2}$	$\frac{1}{\mu} + \frac{\pi_0}{s\mu(1-\rho)^2}$	$\frac{\left(\frac{\lambda}{\mu}\right)^s \mu \pi_0}{(s-1)!(s\mu-\lambda)^2}$

■ **Tableau 4 :** Les mesures de performances d'un système M/M/S.

1.8.3 Système M/M/∞

Pour ce système on a une infinité de serveurs identiques et indépendants les uns des autres du coup il ne permet en aucun cas la création d'une file d'attente c'est-à-dire tout client qui arrive est immédiatement servi. Ce système d'attente est aussi markovien comme les systèmes précédents, le flux des arrivées est poissonnien de paramètre λ ; les durées de services indépendantes suivent la même distribution de paramètre μ .

Pour le processus de naissance et de mort associé à ce système on a :

$\lambda_n = \lambda$	$\forall n \geq 0$
$\mu_n = n\mu$	$\forall n \geq 1$

Le graphe correspondant à ce modèle est :

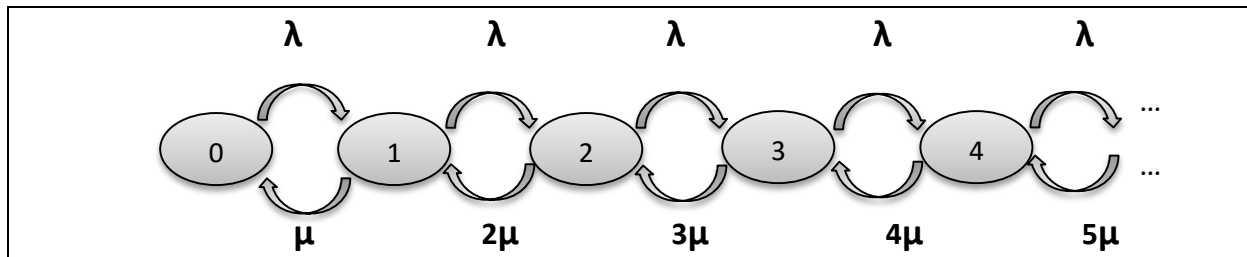


Figure 8 : Graphe de transition d'un système d'attente M/M/∞.

Le régime stationnaire :

On avait déjà mentionné que la capacité de la station de service est infinie, donc la condition de stabilité est toujours satisfaite.

Distribution stationnaire du système M/M/∞ :

Soit π_n la probabilité stationnaire de n client dans le système. A l'aide de la méthode de la résolution de l'équation de Kolmogorov pour un processus de naissance et de mort on trouve :

$$\pi_n = \pi_0 \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} = \pi_0 \prod_{i=1}^n \frac{\lambda}{i\mu} = \pi_0 \frac{\lambda^n}{n!\mu^n} = \pi_0 \frac{\rho^n}{n!} \quad \left(\rho = \frac{\lambda}{\mu}\right)$$

On a $\pi_0 = e^{-\rho}$

En remplaçant dans la formule précédente, π_0 par sa valeur on trouve le résultat final suivant:

$$\pi_n = \frac{\rho^n}{n!} e^{-\rho}$$

Les mesures de performance du système M/M/∞

Nombre moyen de clients dans le système :

$$L = \sum_{n=0}^{\infty} n \pi_n = \sum_{n=0}^{\infty} n \frac{\rho^n}{n!} e^{-\rho} = \rho$$

Temps moyen de service d'un client dans le système :

$$W = \frac{L}{\lambda} = \frac{1}{\mu}$$

L	W
$\frac{\lambda}{\mu}$	$\frac{1}{\mu}$

■ **Tableau 5** : Les mesures de performances d'un système M/M/∞.

Remarque :

Le nombre moyen **L_q** de clients dans la file et la durée moyenne **W_q** d'attente d'un client sont nuls car la file d'attente n'existe pas, donc forcément il n'y aura pas une durée d'attente.

1.8.4 Système d'attente M/M/1/K :

La file M/M/1/K est une file markovienne identique à la file M/M/1 mise à part que le nombre maximum de client admissible simultanément K de cette dernière est fini c'est-à-dire un client qui arrive dans le système et trouve devant lui K-1 autres clients en attente sera automatiquement rejeté à savoir que ce système ne contient qu'un seul serveur. Elle n'a donc aucune condition de stationnarité quel que soit l'intensité du trafic vue que la capacité finie évite l'explosion de la file. Le processus des arrivées est poissonnien de taux λ , la distribution du temps de service est exponentielle de paramètre μ et la discipline d'attente est FIFO (first in

first out). Le nombre de client k dans le système étant fini, on notera alors l'ensemble des états par $E = (1, 2, 3, \dots, K)$.

Ce processus est un processus de naissance et de mort dont les taux de transitions sont :

$$\lambda_n = \begin{cases} \lambda & \text{si } n \leq K \\ 0 & \text{si } n > K \end{cases}$$

$$\mu_n = \mu \quad \forall n \geq 1$$

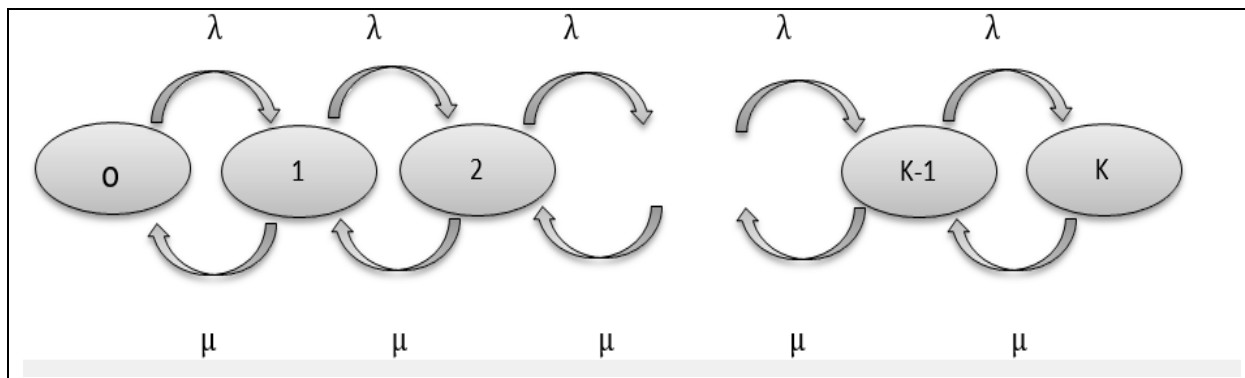


Figure 9 : le graphe de transition d'un système M/M/1/K

Distribution stationnaire du système M/M/1 /K :

A l'aide de la procédure des équations de Chapman-Kolmogorov par déduction du diagramme on obtient les équations suivantes :

Les équations de Kolmogorov sont :

$$\begin{cases} \lambda \pi_0 = \mu \pi_1 \\ (\lambda + \mu) \pi_n = \lambda \pi_{n-1} + \mu \pi_{n+1} & \text{pour } n = 1, 2, 3, \dots, K - 1 \\ \lambda \pi_{K-1} = \mu \pi_K \\ \sum_{n=0}^{\infty} \pi_n = 1 \end{cases}$$

Où π_n , $n=1, 2, 3, \dots$ représente la probabilité qu'il y ait n clients dans le système.

La résolution du système ci-dessus donne la distribution stationnaire suivante :

$$\pi_n = \begin{cases} \frac{1-\rho}{1-\rho^{k+1}} \rho^n & \text{si } \rho \neq 1 \\ \frac{1}{k+1} \rho^n & \text{si } \rho = 1 \end{cases} \quad \text{Avec} \quad \pi_0 = \begin{cases} \frac{1-\rho}{1-\rho^{k+1}} & \text{si } \rho \neq 1 \\ \frac{1}{k+1} & \text{si } \rho = 1 \end{cases}$$

Mesures de performances du système M/M/1/K :

Le nombre moyen de clients dans le système :

$$L = \sum_{n=0}^K n \pi_n \quad \text{Si } \rho \neq 1$$

$$L = \sum_{n=1}^K n \frac{(1-\rho)}{1-\rho^{k+1}} \rho^n = \frac{(1-\rho)}{1-\rho^{k+1}} \sum_{n=1}^K n \rho^n = \frac{(1-\rho)}{1-\rho^{k+1}} \rho \sum_{n=1}^K n \rho^{n-1} \quad \text{Si } \rho \neq 1$$

D'où

$$L = \frac{\rho}{(1-\rho)(1-\rho^{k+1})} [1 - (K+1)\rho^k + K\rho^{k+1}] \quad \text{Si } \rho \neq 1$$

Si $\rho = 1$

$$L = \sum_{n=0}^K n \pi_n = \sum_{n=0}^K \frac{n}{k+1} = \frac{1}{k+1} \sum_{n=0}^K n = \frac{1}{k+1} \frac{K(K+1)}{2}$$

D'où

$$L = \frac{K}{2}$$

En utilisant les formules de Little, on obtient les autres paramètres.

λ' est le taux de client qui rentre effectivement dans le système .Si on a K clients dans le système, les arrivées seront rejetés avec un taux λP_K d'où le taux d'arrivé qui rentre dans le système est :

$$\begin{aligned} \lambda' &= \lambda - \lambda P_K \\ &= \lambda (1-P_K) \end{aligned}$$

Par conséquent on a :

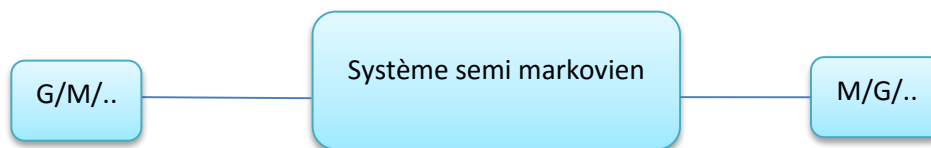
L	Lq	W	Wq
$\frac{\rho}{(1-\rho)(1-\rho^{K+1})} [1 - (K + 1)\rho^K + K\rho^{K+1}]$ <p style="text-align: center;">si $\rho \neq 1$</p> $\frac{K}{2}$ <p style="text-align: center;">si $\rho = 1$</p>	$L + \frac{\lambda(1-P_K)}{\mu}$	$\frac{Lq}{\lambda(1-P_K)}$	$\frac{L}{\lambda(1-P_K)}$

■ **Tableau 6** : Les mesures de performances d'un système M/M/1/K.

1.9 Description d'un système semi markovien :

L'absence de l'exponentialité c'est-à-dire lorsque l'on s'écarte de l'hypothèse d'exponentialité de l'une des deux quantités stochastiques (le temps des inter-arrivées et la durée de service) ou en prenant en compte certaines spécificités des problèmes par l'introduction des paramètres supplémentaires, on aboutit à un modèle semi markovien.

On distingue ainsi deux types de modèle semi markovien représentés par la figure ci-dessous.



✚ **Figure 10** : les systèmes semi markoviens

Dans cette partie, on s'intéressera donc à l'étude des deux systèmes élémentaires du modèle semi markovien à savoir le M/G/1 et le G/M/1.

1.9.1 Système M/G/1 :

Le système M/G/1 est un système semi markovien dont le processus d'arrivée des clients dans le système est toujours poissonien de taux λ , le temps de service d'un client est distribué selon une loi générale différente de la loi exponentielle dont la fonction de répartition est $B(t)$

de densité $b(t)$. Tout comme le système M/M/1, le système M/G/1 ne comporte qu'un seul serveur et une file FIFO de capacité illimitée [15].

Considérons le processus $\{(X_n)_{n \geq 1} = (X(t_n))_{n \geq 1}, t_n > 0\}$ où $X(t_n)$ est le nombre de client dans le système aux différents instants de départ du $n^{\text{ième}}$ client t_n , avec $n=1, 2, 3, \dots$

Le processus $(X_n)_{n \geq 1}$ est une chaîne de Markov ergodique et à temps discret. Cette assertion peut être facilement vérifiée de la manière suivante :

$$X_{n+1} = \begin{cases} X_n - 1 + A_{n+1} & \text{si } X_n \geq 1 \\ A_{n+1} & \text{si } X_n = 0 \end{cases}$$

Où :

- X_n est le nombre de client dans le système au $n^{\text{ième}}$ départ
- A_n est le nombre d'arrivées pendant le temps de service du $n^{\text{ième}}$ client.

Les variables aléatoires A_n sont indépendantes entre elles. Soit a_n la distribution commune des variables aléatoires A_n c'est à dire la probabilité que n clients arrivent pendant un temps de service.

On a :

$$P(A_n = k) = a_k = \int_0^{+\infty} P(k \text{ arrivées pendant } t) b(t) dt = \int_0^{+\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} b(t) dt$$

avec $k=1, 2, 3, \dots$

Déterminons la matrice de transition $P = (a_{i,j})$ de la chaîne de Markov $(X_n)_{n \geq 1}$ dite chaîne de Markov incluse du processus $\{(X_t)_{t \geq 0}\}$.

On pose $P_{i,j} = P(X_{n+1} = j / X_n = i)$ la probabilité de transition de l'état i vers l'état j , telle que :

$$\begin{cases} P_{0,j} = a_j & \text{si } j \geq 0 \\ P_{i,j} = a_{j-1} + 1 & \text{si } 1 \leq i \leq j + 1 \\ P_{i,j} = 0 & \text{sinon} \end{cases}$$

D'où la lecture se fait de la manière suivante :

- si $X_n = 0$, et $[X_{n+1} = j]$ ceci correspond à l'arrivée de j clients pendant le service du client $n+1$.
- si $X_n = i$, on a $[X_{n+1} = j]$ si $X_{n+1} - X_n = j-i$, ce qui correspond à l'arrivée de $j-i+1$ clients pendant le service du client i .

La matrice de transition de la chaîne de Markov incluse est donc donnée par :

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & \dots \\ a_0 & a_1 & a_2 & \dots & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & a_2 \\ 0 & 0 & 0 & a_0 & a_1 \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Cette chaîne de Markov est **irréductible** car le passage d'un état vers n'importe quel autre état est possible d'après la matrice donc on peut facilement montrer qu'elle converge vers une distribution limite si le système est stable.

L'étude de régime stationnaire :

Ce système est stable que si $\rho < 1$ avec $\rho = \lambda E(s)$ où λ est le taux d'arrivée et $E(s)$ la moyenne du temps de service.

Distribution stationnaire :

On suppose que le régime stationnaire est vérifié puis on calculera la fonction génératrice $\pi(z)$ de la distribution stationnaire de la chaîne de Markov induite

$\pi = (\pi_0, \pi_1, \pi_2, \dots)$ vue qu'en général son calcul ne sera pas possible.

$$\pi = \pi P$$

$$\text{Où : } \pi_j = \sum_{i=0}^{+\infty} \pi_i P_{i,j} = a_j \pi_0 + \sum_{i=0}^{j+1} a_{j-i+1} \pi_i - a_{j+1} \pi_0 \quad (j = 1, 2, 3, \dots)$$

En introduisant les fonctions génératrices suivantes :

$$\pi(z) = \sum_{j=0}^{+\infty} \pi_j z^j ; \quad A(z) = \sum_{i=0}^{+\infty} a_i z^i ; \quad \text{et} \quad C(z) = \sum_{j=0}^{+\infty} c_j z^j = \pi(z) A(z) \quad \text{avec}$$

$$A = (a_0, a_1, a_2, \dots) ; \quad c_{j+1} = \sum_{i=0}^{j+1} \pi_i a_{j-i+1} ;$$

On obtient alors:

$$\pi(z) = \frac{\pi_0 A(z)(z-1)}{z-A(z)} \quad \text{et} \quad \pi_0 = 1 - \rho$$

Mesures de performance du système M/G/1 :

Le nombre moyen de clients dans le système : (Formules de Pollaczek Khinchine) :

Le nombre moyen de clients dans une file M/G/1 avec des arrivées Poissonniennes de taux λ et une durée de service s de loi quelconque (admettant au moins une espérance et une variance) est

$$L = \rho + \frac{\rho^2 + \lambda^2 \sigma s^2}{2(1-\rho)} \quad ; \quad \text{avec} \quad \rho = \lambda E(s)$$

En utilisant les formules de Little, on obtient les autres paramètres

L	Lq	W	Wq
$\rho + \frac{\rho^2 + \lambda^2 \sigma s^2}{2(1-\rho)}$	$\frac{\rho^2 + \lambda^2 \sigma s^2}{2(1-\rho)}$	$\frac{\rho^2 + \lambda^2 \sigma s^2}{2\lambda(1-\rho)}$	$\frac{\rho^2 + \lambda^2 \sigma s^2}{2\lambda(1-\rho)}$

■ Tableau 7 : Les mesures de performances d'un système M/G/1.

Où $E(s)$ et σ^2 désignent respectivement la moyenne et la variance du temps de service.

1.9.2 Système G/M/1

Le système G/M/1 est un système semi markovien dont le processus des arrivées des clients dans le système suit une loi générale différente de la loi exponentielle, le temps de service est exponentiellement distribué, la capacité de la file ainsi que celle de la source sont infinies et le service est fourni par un unique serveur. Le processus $\{X(t), t \geq 0\}$ où $X(t)$ désigne le nombre de client dans le système à la date t n'est pas markovien [15].

Considérons le processus $\{(X_n)_{n \geq 1} = (X(t_n))_{n \geq 1}, t_n > 0\}$ où $X(t_n)$ est le nombre de client dans le système à l'instant qui précède immédiatement l'arrivée du $n^{\text{ième}}$ client est une chaîne de Markov. on a :

$$X_{n+1} = X_n + 1 - \Delta_{n+1} \quad \text{Où } \Delta_{n+1} \text{ est le nombre de client servit entre l'arrivée } C_n \text{ et } C_{n+1}.$$

Soit $P_{i,j} = 0 \dots \dots \dots P(X_{n+1} = j / X_n = i)$ la probabilité pour que dans l'intervalle $[t_n, t_{n+1}]$ il y ait exactement $i+1-j$ départ.

$$\begin{cases} P_{i,j} = 0 & \text{si } j > i + 1 \\ P_{0,j} = \int_0^{+\infty} (1 - e^{-\mu x})^{1-j} (e^{-\mu j x}) dA(x) & \text{si } 1 \geq j \\ P_{i,j} = \int_0^{+\infty} \frac{(\mu x)^{i+1-j}}{(i+1-j)!} e^{-\mu x} dA(x) & \text{si } 1 < j < i + 1, \quad i \geq 1 \end{cases}$$

Où $A(x)$ est la distribution des arrivées

A partir de cette matrice de transition $P = \|P_{i,j}\|$ on peut donc obtenir la distribution stationnaire.

Distribution stationnaire :

La distribution stationnaire $\{\pi_k\}$ de la chaîne de la chaîne de Markov $\{x_n\}$ est solution du système d'équation algébrique $\pi = \pi P$. Si $\lambda < \mu$, alors cette distribution pour le système G/M/1 est géométrique.

$$\pi_k = (1 - \sigma)\sigma^k \quad (k = 0, 1, 2, \dots)$$

Où σ est l'unique solution de l'équation $\sigma = \hat{A}(\mu - \mu\sigma)$ dans le domaine $0 < \sigma < 1$.

Les mesures de performances :

Le nombre moyen de clients dans la file d'attente : il est donné par :

$$Lq = \frac{\sigma\rho}{1-\sigma}$$

Le temps moyen d'attente dans la file :

$$Wq = \frac{\sigma E(s)}{1-\sigma} \quad \text{où} \quad E(s) = \frac{1}{\mu}$$

En utilisant les formules de Little, on obtient les autres paramètres.

L	Lq	W	Wq
$\frac{\rho}{1-\sigma}$	$\frac{\sigma\rho}{1-\sigma}$	$\frac{E(s)}{1-\sigma}$	$\frac{\sigma E(s)}{1-\sigma}$

■ **Tableau 8** : Les mesures de performances d'un système G/M/1.

Remarque :

A partir de la distribution stationnaire obtenue pour le système G/M/1 on peut donc déduire celle du système M/M/1. Pour cela, on considérera la distribution des arrivées $A(x)$ exponentielle de paramètre λ et la transformée de Laplace de cette distribution correspond à :

$$\hat{A}(s) = \frac{\lambda}{\lambda+s}.$$

Conclusion

L'intérêt de ce chapitre a été porté sur l'étude des systèmes d'attente classiques (markoviens et semi-markoviens) et leurs caractéristiques en passant par une introduction de quelques notions de base sur la théorie de files d'attente.

Le service d'un client dans les systèmes de file d'attente ne s'effectue pas toujours de manière continue il peut donc y avoir des arrêts momentanés de service. A cet effet il faut donc considérer les systèmes d'attente avec interruption de service.

2 Interruption de service dans les systèmes d'attente

Introduction :

Dans la théorie des files d'attente classiques, il est supposé qu'un client dès son accès au serveur il effectue son service et ceci de façon continue puis quitte le système. Les systèmes de files d'attente complexes tentent de prendre en considération des phénomènes qui empêchent la continuité du service. Un tel système est connu sous le nom de « système de file d'attente avec interruption de service ». Son étude est motivée par diverses applications notamment dans les réseaux de télécommunication, dans la modélisation de la politique de traitement dans les centres de santé etc.

L'influence de ces phénomènes ne peut être étudiée qu'en spécifiant l'interaction entre le processus d'interruption et le processus de service. En effet cette interaction peut ne pas intervenir pendant le service du client dans ce cas elle est dite non préemptive, ou contrairement elle est donc qualifiée préemptive.

Dans un système d'attente, l'interruption de service d'un client peut être causée par divers facteurs on peut citer entre autres : les arrivées négatives, les vacances du serveur, les pannes du serveur etc.

Nous allons donc aborder dans ce chapitre quelques facteurs d'interruption de service portant sur les systèmes d'attente.

2.1 Système de file d'attente avec vacance de serveur :

L'étude de la période d'indisponibilité momentanée du serveur dans un système d'attente ou encore connu sous le nom de vacance de serveur, rend les modèles de files d'attente plus réalistes et plus flexibles [11].

Elle a été premièrement abordée par Levy et Yechiali puis développée plus tard par d'autres chercheurs. En effet dans les systèmes d'attente, les vacances de serveurs peuvent résulter de divers facteurs.

La maintenance des machines dans un processus de production peut être considérée comme un exemple de vacance simple.

Ces modèles ont les mêmes caractéristiques que les modèles précédemment étudiés dans le chapitre 1 mais ces derniers se distinguent par une possibilité de vacance du serveur.

Ainsi un système de file d'attente avec vacance est graphiquement représenté par le schéma ci-dessous.

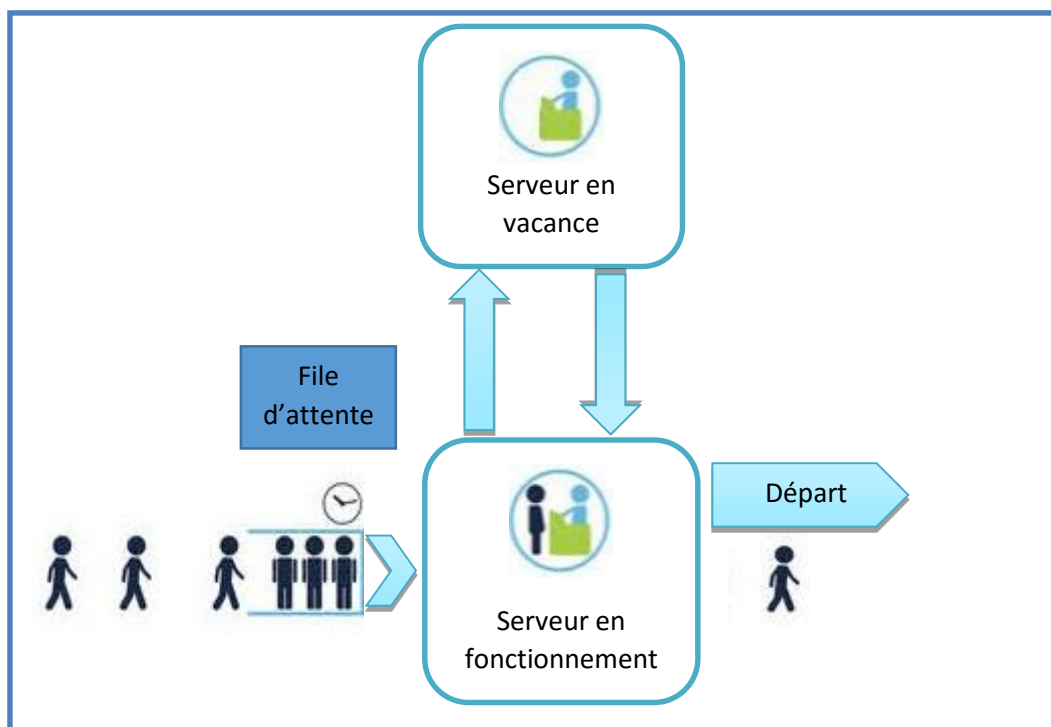


Figure 11 : schéma représentatif d'un système d'attente avec vacance du serveur

2.1.1 Les caractéristiques des politiques de vacances

Les aspects caractérisant la politique de vacance sont :

✓ Règle de début de vacance :

Il existe deux cas de vacances du serveur, le premier cas celui d'un service non- exhaustif et le deuxième cas celui d'un service exhaustif.

• Vacance d'un service non-exhaustif :

Ce cas se caractérise par une possibilité de vacance du serveur même s'il y'a des clients dans le système. On peut notamment citer quelques cas de vacance non-exhaustive.

○ Service avec barrière :

Dans le cas du service avec barrière, le service n'est accordé qu'aux clients qui étaient dans la file d'attente au moment du retour de la vacance du serveur c'est-à-dire le serveur prend une autre vacance que si le nombre de client en file d'attente à son retour de vacance s'épuise (les clients sont tous servis).

○ Service limité :

Dans un système avec service limité, après une vacance, le serveur n'est apte de prendre une autre vacance que lorsqu'il est vide ou s'il atteint un nombre k préalablement fixé c'est à dire avant la prise d'une autre vacance par le serveur, il doit cependant servir au plus k clients.

• Vacance d'un service exhaustif

Pour le système de file d'attente avec vacance d'un serveur exhaustif, le serveur prend exactement une vacance dès que le système est vide. S'il trouve un système vide à son retour de vacance, il devient inactif jusqu'à l'arrivée d'un client.

✓ Règle de fin de vacance :

Cette règle caractérise la reprise du service par le serveur après une vacance. On distingue entre autre deux grandes politiques populaires qui sont :

1. Une vacance multiple

Celle-ci représente le cas d'une prise successive des vacances par le serveur jusqu'à ce qu'il trouve au moins un client en attente à la fin d'une vacance.

2. Un service avec vacance unique

Contrairement au premier cas, le serveur n'est apte de prendre une autre vacance que lorsqu'il sert au moins un client après son retour de la précédente .toutefois, s'il n'y a pas de client présent dans la file alors le serveur reste en inactif.

✓ Durée d'une vacance :

Dans la plupart des temps on suppose que les vacances du serveur sont des variables aléatoires indépendantes et identiquement distribuées avec une certaine loi de probabilité selon les caractéristiques du système modélisé. Ainsi pour la détermination de la durée de la vacance, deux politiques sont indispensables à savoir :

◆ La politique N-vacance (N-Policy) :

Pour cette politique, le serveur ne reprend service que si un nombre maximum N de clients dans la file est atteint.

◆ La politique T-vacance

Dans ce cas-ci, la reprise du service est conditionnée à l'écoulement d'un temps T c'est-à-dire le serveur prend une vacance d'une durée T et une fois cette durée épuisée il reprend son fonctionnement.

2.1.2 Notations

La séquence de symboles utilisée pour représenter les files d'attente avec vacance est de la forme suivante:

$$A/B/C/D/E/F_V$$

Telle que :

- A** : Un code décrivant le processus d'arrivée.
- B** : décrit la distribution des temps de service d'un client.
- C** : le nombre de serveur.
- D** : La capacité du système.
- E** : la capacité de la source.
- F_v**: La discipline de service.

Où F_v correspond à :

$$\left\{ \begin{array}{l} E: \text{service exhaustif.} \\ L: \text{service limité.} \\ b: \text{service avec barrière.} \\ B: \text{service de bernoulli.} \\ N: \text{service } n - \text{vacance.} \end{array} \right.$$

Et

v peut être $\left\{ \begin{array}{l} U: \text{vacance unique.} \\ M: \text{vacance multiple.} \\ H: \text{vacance hybride.} \end{array} \right.$

Nous allons donc faire l'étude du système M/M/1 en guise d'exemple pour la discipline de service exhaustif et vacance simple de serveur.

2.1.3 Modèle M/M/1/∞/∞/Eu: [11]

Description mathématique

Nous considérons un système de file d'attente dans lequel les clients arrivent suivant un processus de poisson de taux λ . Le service est fourni par un unique serveur de capacité illimitée. Le temps de service et les temps de vacances sont distribués de manière exponentielle de débit respective μ et γ . On suppose qu'à l'arrivée d'un client dans le système celui-ci s'impatiente peu importe l'état du système. On suppose également que la durée de cette impatience suit une loi exponentielle de taux ϵ .

On note les dimensionnes $C(t)$, $N(t)$ de la chaîne de Markov représentant ce système, avec l'espace d'états $S= \{(c, n) : c=0,1 \quad n=1, 2,3,\dots\}$ où $N(t)$ désigne le nombre total de clients dans le système et $C(t)$ indique l'état du serveur au temps t .

$$C(t)=\begin{cases} 0 & \text{si le serveur est en vacance} \\ 1 & \text{s'il sert un client ou est inactif} \end{cases}$$

Soit $P_{c,n}=P(C(t) = c, N(t) = n) \quad c = 0,1 ; \quad n = 0$ définissant les probabilités stationnaires de la chaîne de Markov.

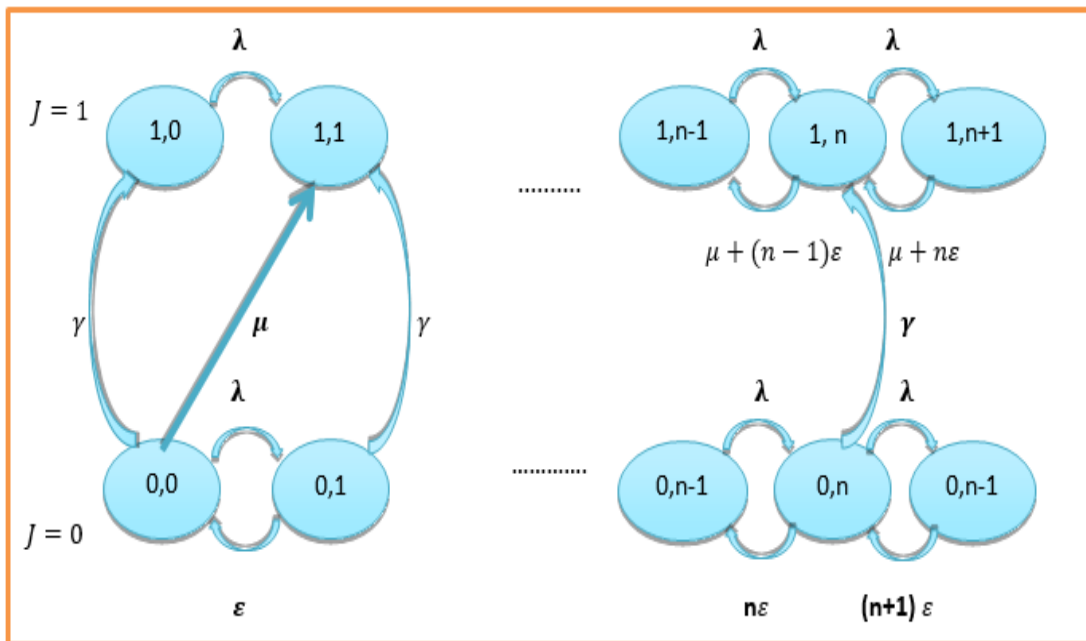


Figure 12 : Graphe de transition d'un système d'attente avec vacance

A l'aide de la méthode des équations de Chapman-Kolmogorov, nous aboutissons aux équations d'équilibre suivantes :

$$(\lambda + \gamma)P_{0,0} = \epsilon P_{0,1} + \mu P_{1,1} \tag{2.1}$$

$$(\lambda + \gamma + n\epsilon)P_{0,n} = \lambda P_{0,n-1} + (n + 1)\epsilon P_{0,n+1} \tag{2.2}$$

$$\lambda P_{1,0} = \gamma P_{0,0} \tag{2.3}$$

$$(\lambda + \mu + (n - 1)\epsilon)P_{1,n} = \lambda P_{1,n-1} + \gamma P_{0,n} + (\mu + n\epsilon)P_{1,n+1} \tag{2.4}$$

On distinguera donc ces cas particuliers grâce aux valeurs prises par certains paramètres de ce système :

Si :

- $\varepsilon = 0$ le modèle se résumera donc à un modèle d'attente M/M/1 avec vacance simple et clients patients.
- $\varepsilon = 0$ uniquement dans l'équation (2.4) alors on obtiendra un M/M/1 avec vacance simple dans lequel l'impatience du client ne se manifeste que si le serveur est en vacance.
- $\gamma = 0$ ce cas représentera alors un système M/M/1 sans vacance avec des phénomènes d'impatiences.

En utilisant les équations obtenues grâce à la méthode de Chapman Kolmogorov tout en supposant que $\frac{\lambda}{\mu} < 1$ et $\frac{\varepsilon}{\mu} < 1$, on exprime les fonctions génératrices partielles en terme de $P_{0,0}$ comme suit:

$$P_0(z) = P_{0,0} \exp\left(\frac{\lambda}{\varepsilon} z\right) (1 - z)^{-\frac{\gamma}{\varepsilon}} \left[1 - \frac{A(z)}{A(1)} \right] \quad (2.5)$$

$$P_1(z) = P_{0,0} \exp\left(-\frac{\lambda}{\varepsilon} z\right) z^{\frac{\mu}{\varepsilon}-1} \left[-\frac{\gamma}{\varepsilon} B(z) + \frac{D(z)}{A(1)} + \frac{(\mu-\varepsilon)\gamma}{\varepsilon\lambda} C(z) \right] \quad (2.6)$$

Où

$$A(z) = \int_0^z \exp\left(-\frac{\lambda}{\varepsilon} s\right) (1 - s)^{\frac{\gamma}{\varepsilon}-1} ds \quad (2.7)$$

$$B(z) = \int_0^z s^{\frac{\mu}{\varepsilon}-1} (1 - s)^{-\left(\frac{\gamma}{\varepsilon}+1\right)} \left(1 - \frac{A(s)}{A(1)} \right) ds \quad (2.8)$$

$$C(z) = \int_0^z \exp\left(-\frac{\lambda}{\varepsilon} s\right) s^{\frac{\mu}{\varepsilon}-2} ds \quad (2.9)$$

$$D(z) = \int_0^z \exp\left(-\frac{\lambda}{\varepsilon} s\right) s^{\frac{\mu}{\varepsilon}-1} (1 - s)^{-1} ds \quad (2.10)$$

Ainsi on obtiendra par la suite des calculs l'expression de $P_{0,0}$ qui est de :

$$P_{0,0}(z) = \left[\frac{\varepsilon}{\gamma A(1)} + e^{\frac{\lambda}{\varepsilon}} \left(-\frac{\gamma}{\varepsilon} B(1) + \frac{D(1)}{A(1)} + \frac{(\mu - \varepsilon)\gamma}{\varepsilon \lambda} C(1) \right) \right]^{-1} \quad (2.11)$$

Les mesures de performance

➤ Le nombre moyen de clients dans le système :

Soit $E(N_j)$ la moyenne du nombre de client N_j dans le système à l'état j du serveur.

-La moyenne de la taille du système lorsque le serveur est en vacances est :

$$E(N_0) = \sum_{n=0}^{\infty} n P(j = 0, N = n)$$

Et l'on obtient après calcul

$$E(N_0) = \frac{\lambda}{\gamma + \varepsilon} \frac{\varepsilon}{\gamma A(1)} P_{0,0} \quad \text{avec } P_{0,0} \text{ donnée par (2.11)}$$

-La moyenne de la taille du système pendant la période de non vacance est donnée par :

$$E(N_1) = \sum_{n=0}^{\infty} n P(j = 1, N = n)$$

A la fin des calculs on aura :

$$E(N_1) = \frac{\lambda - \mu + \varepsilon}{\varepsilon} + \left[\frac{\lambda}{\lambda + \varepsilon} - \frac{\gamma}{\lambda} + \frac{\mu \gamma}{\varepsilon \lambda} \right] P_{0,0} \quad \text{avec } P_{0,0} \text{ donnée par (2.11)}$$

➤ La durée moyenne de séjour d'un client :

On note par S la durée de séjour totale d'un client ayant terminé ou pas son service dans le système, puis on suppose une durée conditionnelle de séjour $S_{j,n}$ d'un client qui termine son service et ayant trouvé à son arrivé le système à l'état (j,n) . D'où $S_{j,n}$ sera égal à :

$$S_{j,n} = E(S / (X_0 = (j, n+1))) \text{ car le client marqué est dans le système ;}$$

A savoir également que si un client est marqué dès son arrivée dans le système, ce qui correspond à l'état d'inactif du serveur, il aura comme durée totale :

$$E(S_{1,0}) = \frac{1}{\mu}$$

Ainsi par conditionnement de la prochaine transition à une arrivée ou à un départ nous obtenons la relation suivante :

$$E(S_{1,0}) = E(S/(X_0 = (1, n+1)))$$

Et après calcul on obtiendra :

$$E(S_{1,0}) = \frac{n+1}{\mu+n\varepsilon}$$

Les autres mesures de performances seront obtenues à partir de la formule de Little.

2.2 Système d'attente avec panne du serveur :

Sachant que le système parfait n'a pas encore été inventé, dans tous les systèmes d'attente modélisant les phénomènes de la vie courante, on attend des serveurs qu'ils soient en bon état de fonctionnement au moment du service et malheureusement cette disponibilité est affectée par les interruptions de services causée par les pannes aléatoires qui peuvent survenir. Cependant, il existe d'énorme risque dans l'utilisation des serveurs non fiables on peut citer entre autre une augmentation du temps d'attente, une perte de client etc. c'est donc dans cette optique que l'étude des modèles d'attente avec panne des serveurs fut introduits par les chercheurs. On distingue deux types de pannes de serveur qui sont :

2.2.1 Pannes passives

Les pannes passives sont des pannes que le serveur peut avoir juste après les instants de départ d'un client (c'est à dire lorsque le serveur est libre). On le représente par le schéma ci-dessous.

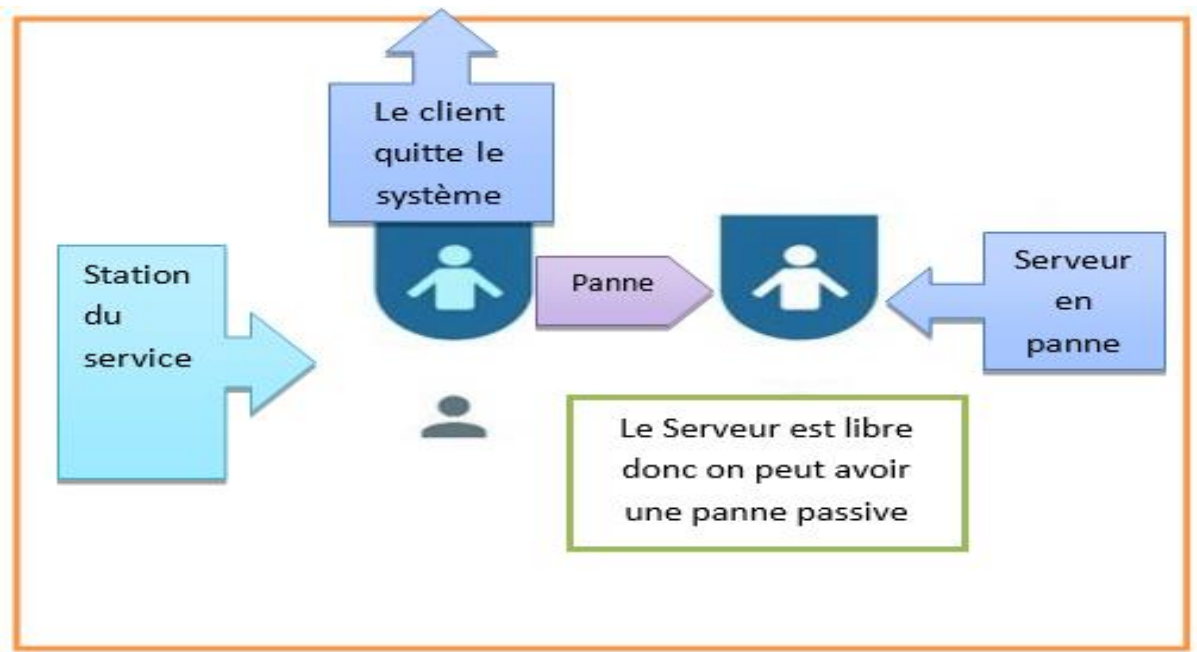


Figure 13 : schéma représentatif d'une panne passive

2.2.2 Pannes actives

Les pannes actives sont également des pannes du serveur qui ne peuvent se produire que lorsque le serveur est occupé par un client (pendant la durée de service d'un client). La représentation schématique du modèle est donnée par :

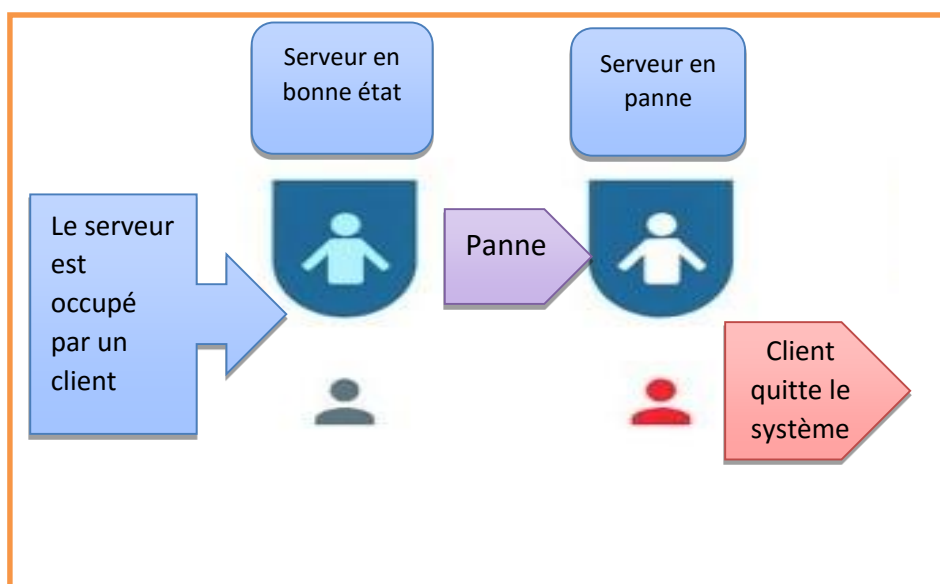


Figure 14 : schéma représentatif d'une panne active

On peut notamment citer quelques cas de pannes actives :

- **Pannes avec perte définitive du client :**

Le client quitte irrémédiablement le système au même instant que survient la panne.

- **Pannes conservatrices :**

Dès que la panne se produit, le client attend la fin de la réparation du serveur puis le service reprend tout en tenant compte de la quantité de service déjà acquise.

- **Pannes non conservatrices :**

Dans ce cas-ci le client reste en attente après une panne de serveur puis après réparation de celui-ci, il prend intégralement son service.

Remarque : Les modèles d'attente avec pannes ont les mêmes caractéristiques que les modèles précédemment étudiés dans le chapitre 1, mais ce qui les distingue, est le risque des pannes. Les pannes constituent dans ce cas un moyen supplémentaire afin qu'un client arrivant quitte le système.

Ainsi d'une façon générale un système d'attente avec panne est décrit par le processus stochastique suivant :

$$S(t) = \{X(t), N(t); t \geq 0\}$$

Où :

- ❑ $N(t)$: est le nombre de clients dans le système à l' instant t .
- ❑ $X(t)$: décrit l'existence de la panne, c'est à dire :
 - ✦ Si : $X(t) = 0$ la panne n'existe pas (Le serveur en bonne état).
 - ✦ Si : $X(t) = 1$ la panne existe (Le serveur en panne).

2.3 Systèmes de files d'attente avec arrivées négatives

Ces modèles se distinguent des deux précédents (modèles avec panne et les modèles avec vacance) car pour ces types, les interruptions du service ne sont pas dues à l'état du serveur

mais plutôt à des arrivées négatives c'est-à-dire en plus des clients positifs qui correspondent à la définition classique des clients, il y a aussi un flux des arrivées négatives dont la présence dans le système n'est pas de recevoir un service mais plutôt d'influencer le système en éliminant les clients positifs.

2.3.1 Le processus d'élimination des clients positifs :

Lorsqu'un client positif arrive au système, si le serveur est libre, il effectue son service ensuite il quitte le système, sinon il rejoint la file d'attente immédiatement, En attendant d'être servi. Par contre, quand le client négatif arrive au système, il élimine immédiatement un ou plusieurs clients positifs de la file. Cette élimination peut être :

- L'élimination individuelle : l'arrivée négative élimine un client positif ;
- L'élimination par groupe : l'arrivée négative élimine un groupe de client du système ;
- Le désastre : un client négatif élimine tous les clients positifs dans le système ;

Et ceci selon les politiques d'élimination suivantes [13] :

- ❑ ECS (élimination du client en service) : le client négatif élimine le client positif qui était en service à son arrivé ;
- ❑ ECF (élimination du client à la fin) : dans ce cas-ci, le client positif qui se trouve à la fin de la file d'attente sera donc éliminé à l'arrivé du client négatif ;
- ❑ EAC (élimination aléatoire d'un client) : cette politique permet d'éliminer le client de façon aléatoire et conjointement avec la discipline du partage de service ;

Remarque : L'arrivée négative dans un système vide est sans effet.

2.3.2 Notation :

La notation abrégée de Kendall basée sur ce type de système est généralement donnée par :

A/B/C/D/politique de service/politique d'élimination

Telle que :

- A** : Un code décrivant le processus d'arrivée.
- B** : décrit la distribution des temps de service d'un client.
- C** : le nombre de serveur.
- D** : la longueur maximale de la file.

2.3.3 Les différents types de clients négatifs [13] :

Depuis l'introduction de ces types de systèmes, plusieurs variétés d'exemples ont été abordées par différents chercheurs afin d'observer les différents comportements de ces systèmes en fonction des différentes politiques (service et élimination). On peut citer entre autre :

- **Clients négatifs de Gelenbe :**

Gelenbe introduit la notion des clients négatifs dans l'étude des systèmes d'attente en se basant sur la politique FIFO pour les arrivées et la politique ECS pour le service.

- **Clients négatifs de boucherie & van Dijk :**

Ces types de systèmes décrivent les systèmes dans lesquels les clients négatifs ainsi que les services peuvent s'accumuler en attendant l'arrivée des clients positifs.

- **Clients négatifs de Henderson :**

Henderson a également modifié ce concept de clients négatifs décrits par boucherie & van Dijk en éliminant l'hypothèse d'accumulation de service c'est-à-dire, Les files d'attentes peuvent contenir un nombre de clients négatifs, mais le serveur arrête le service si aucun client positif n'est présent. Ainsi, les clients négatifs restent dans la file aussi longtemps qu'un client positif ne se présente. Lorsqu'un client positif arrive, l'un des clients négatifs élimine le nouvel arrivant et disparaît aussitôt du système. Les clients négatifs de Henderson ont une patience infinie.

- Clients négatifs actifs :

Les clients négatifs de ce type peuvent s'accumuler et être servis avec des intensités autres que nulles ou infinies. Par exemple, un système de neurones qui subit l'arrivée de doses de stimulant et de tranquillisant.

2.4 Système d'attente avec priorité

La priorité dans les systèmes d'attente est l'un des facteurs d'interruption du service couramment rencontré dans la vie courante car ceci décrit presque en général toutes nos activités quotidiennes comme par exemple, la politique de traitement des patients dans les hôpitaux où une personne inconsciente ou ayant une crise cardiaque bénéficie d'une priorité de service par rapport aux autres qu'ont subi une blessure mineure.

Dans le but d'offrir une qualité de service différente pour différents types de clients, nous contrôlons souvent un système de file d'attente par le mécanisme de priorité. En effet, elle se caractérise par l'arrivée de plusieurs types de clients dans le système, dont certains types sont prioritaires par rapport à d'autres. En considérant donc n comme étant le nombre de classe de clients qui arrive dans le système et que les clients d'une classe i arrivent indépendamment des autres suivant un processus d'arrivée de taux λ_i , $i=1, 2, 3, \dots, n$; alors le taux du flux total des arrivées est donné par $\lambda = \sum_{i=1}^n \lambda_i$. De ce fait la i^{ieme} priorité est accordée aux clients de classe i et elle est supérieure à la k^{ieme} priorité pour tout $k > i$. Les clients de chaque classe disposent donc d'une distribution de service de taux distinct μ_i , $i=1, 2, 3, \dots, n$.

Ainsi la condition d'ergodicité du système est vérifiée si $\sum_{i=1}^n \frac{\lambda_i}{\mu_i} < 1$.

La politique de priorité familière implique une priorité absolue et une priorité relative. [14]

2.4.1 Priorité relative

La priorité relative se caractérise par le fait que l'arrivée d'un client de classe supérieure n n'influe en aucun cas sur le service des clients présents dans les stations de service. En effet, lorsqu'un client jouissant d'une priorité élevée arrive dans le système, il prend une place dans

la file d'attente (s'il le désire bien sûr) et ceci devant tous les clients moins prioritaires présents et attend que le service de tous les clients prioritaires qui le précédaient soit effectué. En outre un client moins prioritaire dans la même situation joint automatiquement la file et devient la dernière personne à être servi.

2.4.2 Priorité absolue

Contrairement au cas précédent, dans le cas d'une priorité absolue un client de classe supérieure a le droit d'interrompre le service d'un client de classe inférieure pour se faire servir, c'est-à-dire un client de haute priorité qui arrive et trouve le serveur occupé par le service d'un autre client de même classe, prend une place dans la file d'attente et ceci devant tous les clients moins prioritaires présents dans cette file, par contre au cas où il le trouve occupé par le service d'un client moins prioritaire, il commence immédiatement son service et le client dont le service a été interrompu par l'arrivée du client prioritaire rejoint la file tout en étant à la tête afin d'être de nouveau pris en charge par le serveur à partir du point où il a été interrompu ou il recommence intégralement son service.

Conclusion:

Ce chapitre a été basé sur une étude analytique des différentes causes d'interruption de service (vacance du serveur, panne du serveur, les arrivées négatives et les priorités) tout en mettant en relief les différentes politiques (d'arrivée et de service), ainsi que leurs effets sur le service d'un client dans les systèmes.

3 L'analyse du système d'attente M/M/1 avec priorité relative et absolue

Introduction

Dans ce chapitre, nous allons effectuer une étude détaillée du système d'attente M/M/1 avec priorité relative et absolue et ceci en passant par la détermination des fonctions génératrices à l'aide des équations de Kolmogorov obtenues grâce aux différents graphes de transition d'état de ce système; puis on déterminera également leurs mesures de performance. Ensuite, nous ferons une illustration numérique de certaines mesures de performance préalablement obtenues.

3.1 Description du Système d'attente M/M/1 avec priorité :

Dans ce système d'attente on considère deux types de clients (clients prioritaires et non prioritaires) qui arrivent au système indépendamment l'un de l'autre suivant des processus de Poisson de taux λ_1 et λ_2 respectivement pour les clients prioritaires et les clients non prioritaires. Ainsi le flux total des arrivées est poissonien de taux $\lambda = \lambda_1 + \lambda_2$. Le service étant fourni par un unique serveur dont la politique de traitement est FIFO. Le temps de service des deux types de clients est exponentiellement distribué de taux μ_1 pour les clients prioritaires et μ_2 pour les non prioritaires. L'admission d'un client (prioritaire ou non prioritaire) au service se fait selon deux possibilités qui dépendent de la nature de la priorité (relative ou absolue).

D'après la description précédente, on peut donc définir l'état du système à l'instant t par le processus stochastique suivant : $\{X(t), t \geq 0\} = \{N(t), M(t), \gamma(t); t \geq 0\}$

Avec :

- $N(t)$: représente le nombre de clients prioritaires dans le système à l'instant t .
- $M(t)$: indique le nombre de clients non prioritaires dans le système à l'instant t .
- $\gamma(t)$: indique le type de client en service $\{0, 1, \text{ou } 2\}$

Où :

$$\gamma(t) = \begin{cases} 0 & \text{si le serveur est libre} \\ 1 & \text{si le serveur est occupé par un client prioritaire} \\ 2 & \text{si le serveur est occupé par un client non prioritaire} \end{cases}$$

Notons alors par $P_{n,m,\gamma}$ la probabilité d'avoir n clients prioritaires, m clients non prioritaires dans le système au temps t ($t \geq 0$), et γ le type de client en service.

On considère que les deux types de clients ont la même durée de service c'est-à-dire $\mu_1 = \mu_2 = \mu$.

L'étude de la stabilité du système

D'après la description précédente de ce modèle, On définit par λ le taux d'arrivée total tel

que :
$$\lambda = \lambda_1 + \lambda_2 \quad \text{et} \quad \rho = \rho_1 + \rho_2 = \frac{\lambda_1}{\mu} + \frac{\lambda_2}{\mu}$$

Et pour que le système soit stable il suffit que : $\rho = \frac{\lambda_1 + \lambda_2}{\mu} < 1$.

Distribution stationnaire

Notre système étant stable ($\rho = \frac{\lambda_1 + \lambda_2}{\mu} < 1$), donc on peut calculer la probabilité $P_{n,m,\gamma}$, mais puisque la résolution de ce système s'avère compliquée, alors on procède donc par la résolution à l'aide des fonctions génératrices, car cette méthode est plus facile par rapport à l'autre méthode (la résolution de l'équation de Kolmogorov).

Soit $G(Z_1, Z_2)$ la fonction génératrice utilisée pour la résolution de ce modèle tel que :

$$G(Z_1, Z_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} Z_1^n Z_2^m P_{n,m,\gamma} = G^1(Z_1, Z_2) + G^2(Z_1, Z_2) + P_{0,0,0}$$

Avec :

$G^1(Z_1, Z_2)$: la fonction génératrice du système lorsque le serveur est occupé par un client prioritaire à l'instant t ($t \geq 0$).

$G^2(Z_1, Z_2)$: la fonction génératrice du système lorsque le serveur est occupé par un client **non** prioritaire à l'instant t ($t \geq 0$).

Et les fonctions génératrices partielles sont introduites comme suites :

$$G^1(Z_1, Z_2) = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} Z_1^n Z_2^m P_{n,m,1} = \sum_{n=1}^{\infty} Z_1^n G_n^1(Z_2)$$

$$G^2(Z_1, Z_2) = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} Z_1^n Z_2^m P_{n,m,2} = \sum_{n=0}^{\infty} Z_1^n G_n^2(Z_2)$$

Où $G_n^2(Z_2)$ et $G_n^1(Z_2)$ s'écrivent sous la forme suivante :

$$G_n^2(Z_2) = \sum_{m=1}^{\infty} Z_2^m P_{n,m,2}$$

$$G_n^1(Z_2) = \sum_{m=0}^{\infty} Z_2^m P_{n,m,1}$$

Dans ce qui suit nous allons donc procéder à l'étude du système décrit ci-dessus tout en spécifiant la nature de la priorité (relative ou absolue).

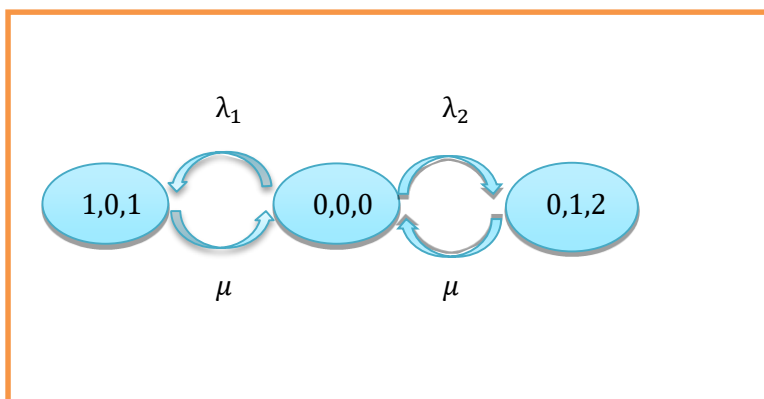
3.1.1 Système d'attente M/M/1 avec Priorité relative

■ Description du système

Lorsqu'un client prioritaire arrive au système et trouve le serveur occupé par un client non prioritaire il rejoint la file et attend jusqu'à ce que ce client non prioritaire termine son service pour être servi, d'autre côté, s'il trouve le serveur occupé par le service d'un client de même type (prioritaire) dans ce cas, il joint automatiquement la file en se plaçant derrière tous les clients prioritaires présents dans cette file et devant tous les clients non prioritaires.

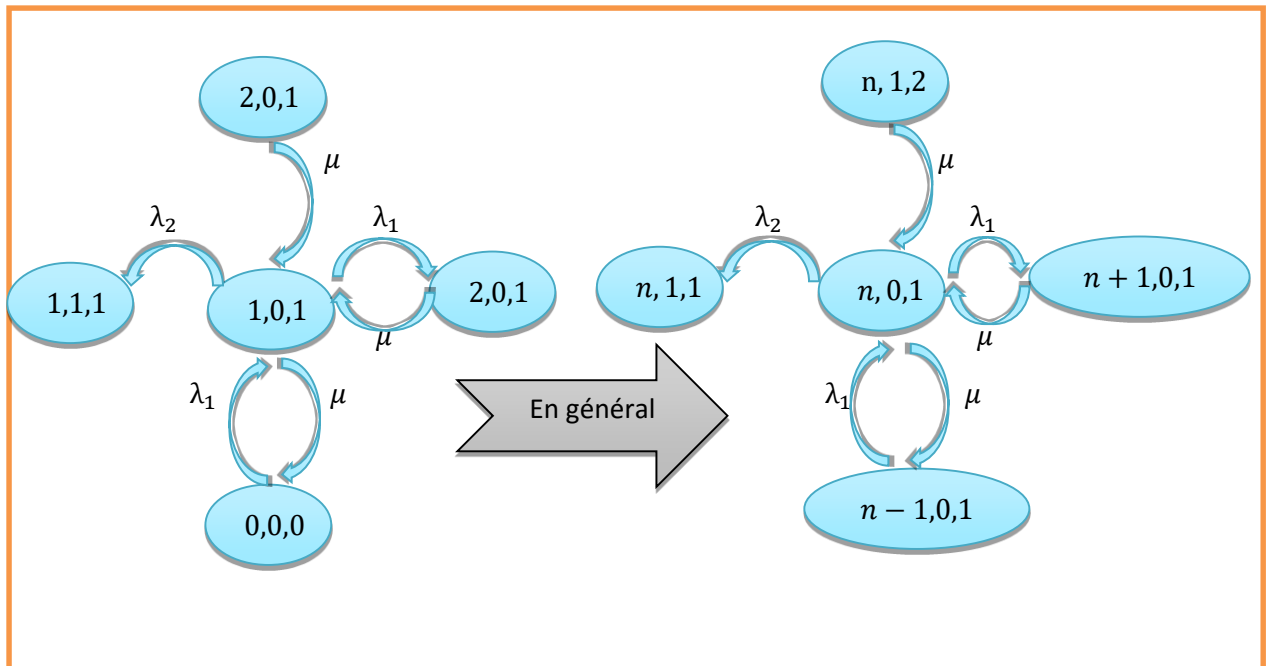
■ Les graphes de transition et les équations de Kolmogorov pour la priorité relative

- **L'état (0, 0,0) :** ce graphe représente l'état (0, 0,0), c'est-à-dire l'état à laquelle le système ne contient aucun client (prioritaire et non prioritaire).



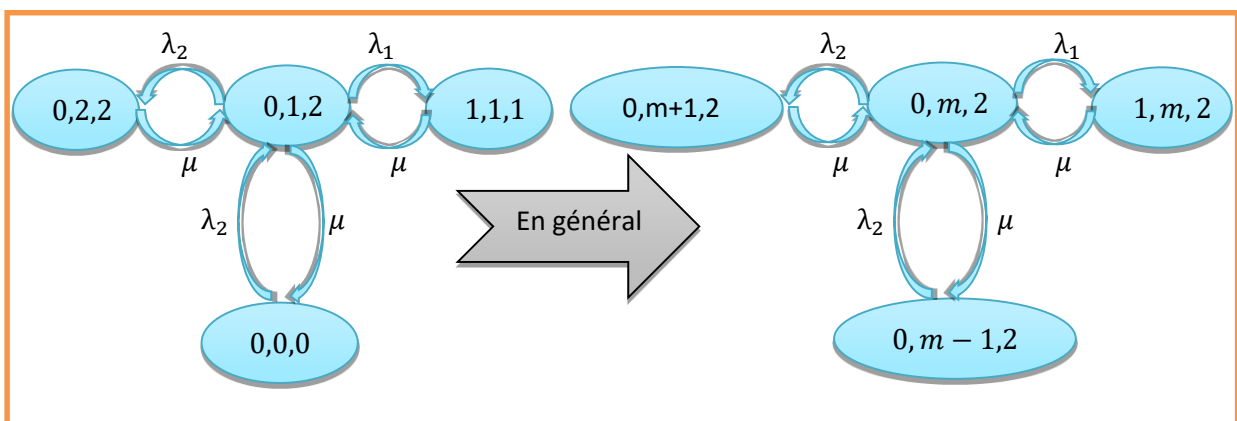
- **L'état (1, 0,1) et (n, 0,1) :** ces deux graphes représentent les deux états (1, 0,1) et (n, 0,1) dont la description se fait de la manière suivante :

Le système ne contient qu'un (resp n) client prioritaire et l'unique serveur est occupé par un client prioritaire.

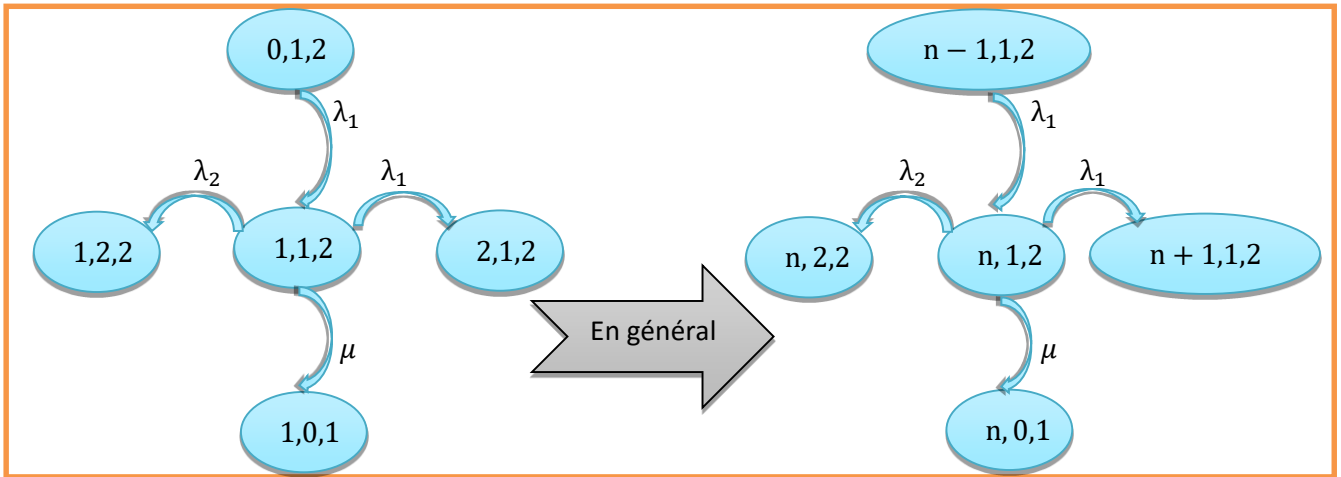


- **L'état (0, 1,2) et (0, m, 2) :** ces deux graphes représentent les deux états (0, 1,2) et (0, m, 2)

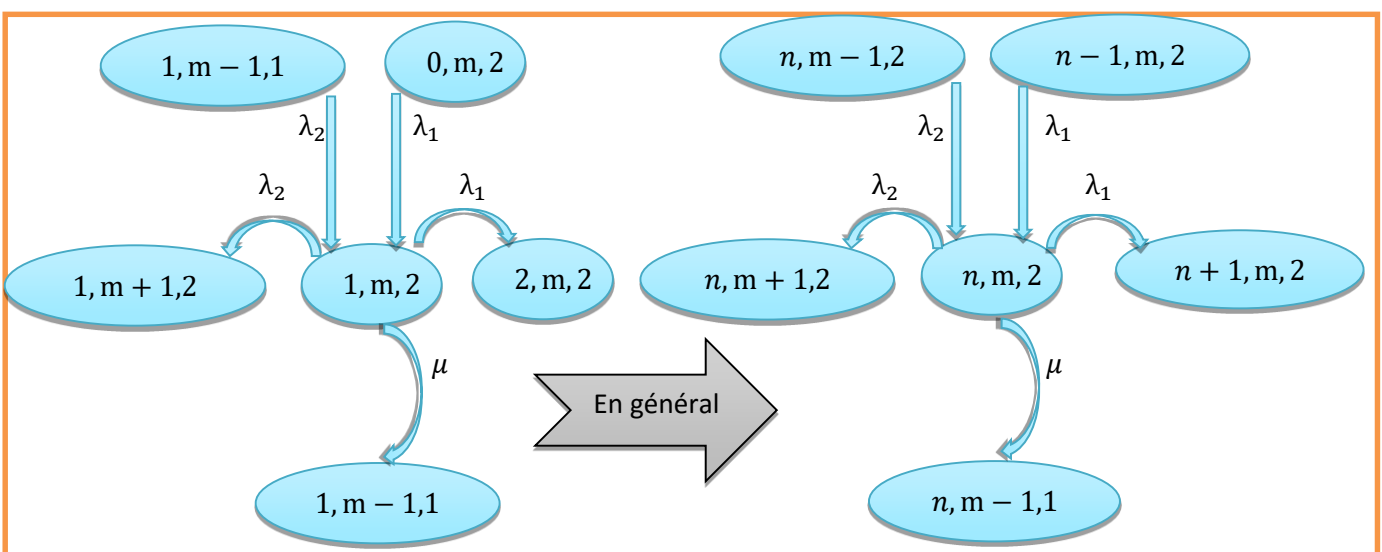
Ces états caractérisent les cas où le système contient 0 client prioritaire, un client (resp m client) non prioritaire pour l'état (0, 1,2) (resp l'état (0, m, 2)) et le serveur étant occupé par un client non prioritaire.



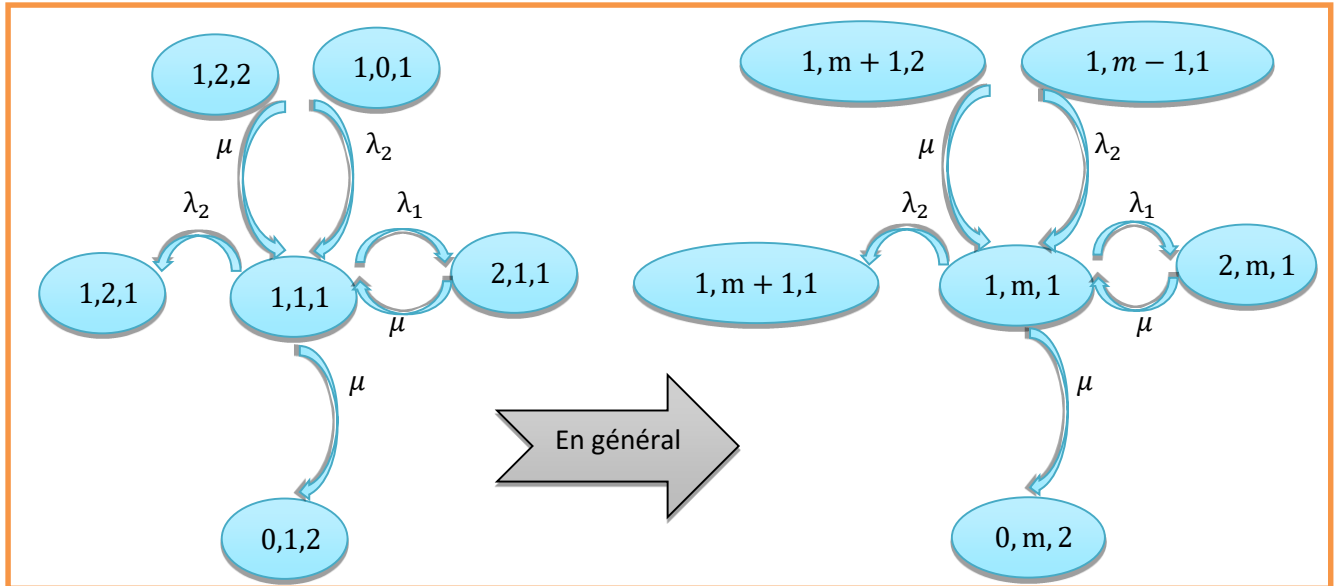
- **L'état (1, 1,2) et (n, 1, 2):** ces deux états représentent le fait que le système contient un (resp n) client prioritaire et le serveur est occupé par l'unique client non prioritaire présent dans ce système. Voici les graphes représentatifs :



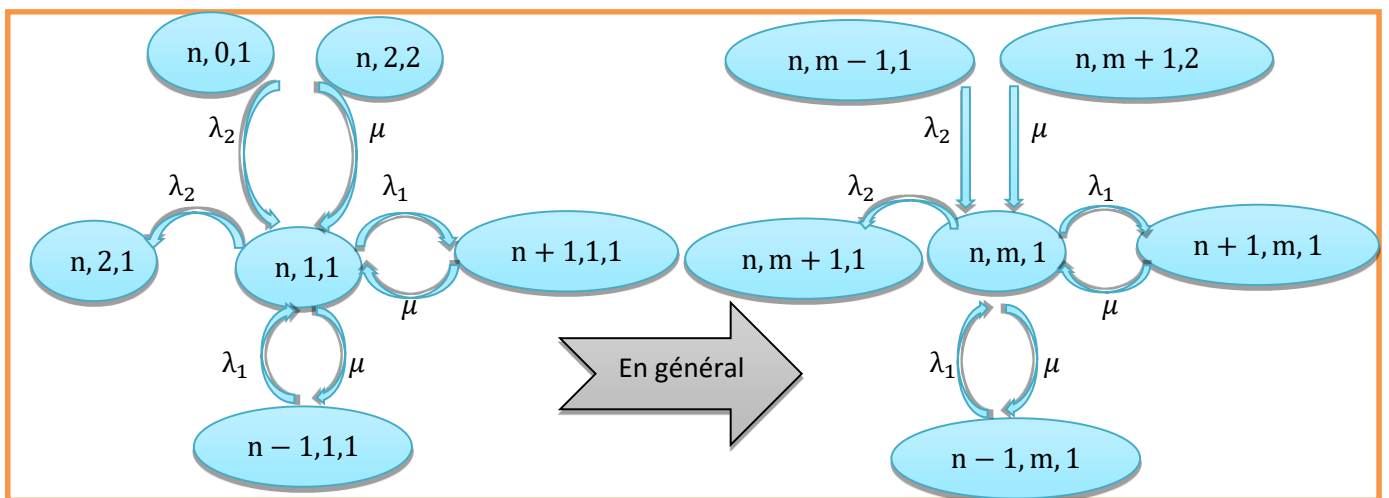
- **L'état (1, m, 2) et (n, m, 2) :** ces deux graphes représentent les deux états (1, m,2) et (n, m, 2) .Pour un (resp n) client prioritaire pour l'état (1, 1,2) (resp (n, 1, 2)) , m client non prioritaire dans le système, et le serveur étant occupé par un client non prioritaire.



- **L'état (1, 1, 1) et (1, m, 1) :** ces graphes ci-dessous représentent les états (1, 1, 1) et (1, m, 1) qui caractérisent le système lorsqu'il contient un client prioritaire, et un (resp m) client non prioritaire, avec un client prioritaire au serveur.



- **L'état (n, 1, 1) et (n, m, 1) :** ces deux graphes ci-dessous représentent les deux états (n, 1, 1) et (n, m, 1) qui correspondent à n client prioritaire, un (resp m) client non prioritaire dans le système, et le serveur étant occupé par un client prioritaire.



A L'aide des graphes de transition précédents nous obtenons le système d'équation (les équations de Kolmogorov) suivant :

$$(\lambda_1 + \lambda_2)P_{0,0,0} = \mu P_{1,0,1} + \mu P_{0,1,2} \quad (1)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{0,1,2} = \lambda_2 P_{0,0,0} + \mu P_{0,2,2} + \mu P_{1,1,1} \quad (2)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{0,m,2} = \lambda_2 P_{0,m-1,2} + \mu P_{0,m+1,2} + \mu P_{1,m,1} \quad (3)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,1,2} = \lambda_1 P_{0,1,2} \quad (4)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,m,2} = \lambda_2 P_{1,m-1,2} + \lambda_1 P_{0,m,2} \quad (5)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,1,2} = \lambda_1 P_{n-1,1,2} \quad (6)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,m,2} = \lambda_2 P_{n,m-1,2} + \lambda_1 P_{n-1,m,2} \quad (7)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,0,1} = \lambda_1 P_{0,0,0} + \mu P_{1,1,2} + \mu P_{2,0,1} \quad (8)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,1,1} = \lambda_2 P_{1,0,1} + \mu P_{2,1,1} + \mu P_{1,2,2} \quad (9)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,m,1} = \lambda_2 P_{1,m-1,1} + \mu P_{2,m,1} + \mu P_{1,m+1,2} \quad (10)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,0,1} = \lambda_1 P_{n-1,0,1} + \mu P_{n,1,2} + \mu P_{n+1,0,1} \quad (11)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,1,1} = \lambda_2 P_{n,0,1} + \lambda_1 P_{n-1,1,1} + \mu P_{n,2,2} + \mu P_{n+1,1,1} \quad (12)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,m,1} = \lambda_2 P_{n,m-1,1} + \lambda_1 P_{n-1,m,1} + \mu P_{n,m+1,2} + \mu P_{n+1,m,1} \quad (13)$$

■ Résolution

On commence la résolution de ce système par l'addition des équations tel que :

{(1) + (2) + (3)} qui donne :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=1}^{\infty} Z_2^m P_{0,m,2} + (\lambda_1 + \lambda_2) P_{0,0,0} = \lambda_2 Z_2 \sum_{m=1}^{\infty} Z_2^m P_{0,m,2} + \frac{\mu}{Z_2} \sum_{m=1}^{\infty} Z_2^m P_{0,m,2} + \mu \sum_{m=0}^{\infty} Z_2^m P_{1,m,1} + \lambda_2 Z_2$$

Alors :

$$(\lambda_1 + \lambda_2(1 - Z_2) + \mu(1 - \frac{1}{Z_2})) G_0^2(Z_2) = \mu G_1^1(Z_2) - (\lambda_1 + \lambda_2(1 - Z_2)) P_{0,0,0} \quad (1^*)$$

{(4) + (5)} qui donne :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=1}^{\infty} Z_2^m P_{1,m,2} = \lambda_2 Z_2 \sum_{m=1}^{\infty} Z_2^m P_{1,m,2} + \lambda_1 \sum_{m=1}^{\infty} Z_2^m P_{0,m,2}$$

Alors :

$$(\lambda_1 + \lambda_2(1 - Z_2) + \mu) G_1^2(Z_2) = \lambda_1 G_0^2(Z_2) \quad (2^*)$$

{(6) + (7)} :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=1}^{\infty} Z_2^m P_{n,m,2} = \lambda_2 Z_2 \sum_{m=1}^{\infty} Z_2^m P_{n,m,2} + \lambda_1 \sum_{m=1}^{\infty} Z_2^m P_{n-1,m,2}$$

Alors :

$$(\lambda_1 + \lambda_2(1 - Z_2) + \mu) G_n^2(Z_2) = \lambda_1 G_{n-1}^2(Z_2) \quad (3^*)$$

{(8) +(9) + (10)} :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=0}^{\infty} Z_2^m P_{1,m,1} = \lambda_2 Z_2 \sum_{m=0}^{\infty} Z_2^m P_{1,m,1} + \mu \sum_{m=0}^{\infty} Z_2^m P_{2,m,1} + \frac{\mu}{Z_2} \sum_{m=0}^{\infty} Z_2^m P_{1,m,2} + \lambda_1 P_{0,0,0}$$

Alors :

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_1^1(Z_2) = \mu G_2^1(Z_2) + \frac{\mu}{Z_2} G_1^2(Z_2) + \lambda_1 P_{0,0,0} \quad (4^*)$$

{(11) +(12) + (13)} :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=0}^{\infty} Z_2^m P_{n,m,1} = \lambda_2 Z_2 \sum_{m=0}^{\infty} Z_2^m P_{n,m,1} + \lambda_1 \sum_{m=0}^{\infty} Z_2^m P_{n-1,m,1} + \frac{\mu}{Z_2} \sum_{m=1}^{\infty} Z_2^m P_{n,m,2} + \mu \sum_{m=0}^{\infty} Z_2^m P_{n+1,m,1}$$

Alors :

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_n^1(Z_2) = \lambda_1 G_{n-1}^1(Z_2) + \mu G_{n+1}^1(Z_2) + \frac{\mu}{Z_2} G_n^2(Z_2) \quad (5^*)$$

A la fin de la somme des équations, on obtient le système des équations suivant :

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu (1 - \frac{1}{Z_2})) G_0^2(Z_2) = \mu G_1^1(Z_2) - (\lambda_1 + \lambda_2(1- Z_2)) P_{0,0,0} \quad (1^*)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_1^2(Z_2) = \lambda_1 G_0^2(Z_2) \quad (2^*)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_n^2(Z_2) = \lambda_1 G_{n-1}^2(Z_2) \quad (3^*)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_1^1(Z_2) = \mu G_2^1(Z_2) + \frac{\mu}{Z_2} G_1^2(Z_2) + \lambda_1 P_{0,0,0} \quad (4^*)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu) G_n^1(Z_2) = \lambda_1 G_{n-1}^1(Z_2) + \mu G_{n+1}^1(Z_2) + \frac{\mu}{Z_2} G_n^2(Z_2) \quad (5^*)$$

Par addition de :

{(2*) et (3*)}, on trouve :

$$(\lambda_1 + \lambda_2(1 - Z_2) + \mu)G^2(Z_1, Z_2) = \lambda_1 Z_1 G^2(Z_1, Z_2) + (\lambda_1 + \lambda_2(1 - Z_2) + \mu)G_0^2(Z_2)$$

$$(\lambda_1(1 - Z_1) + \lambda_2(1 - Z_2) + \mu)G^2(Z_1, Z_2) = (\lambda_1 + \lambda_2(1 - Z_2) + \mu)G_0^2(Z_2)$$

{(4*) et (5*)}, on trouve :

$$\begin{aligned} (\lambda_1 + \lambda_2(1 - Z_2) + \mu)G^1(Z_1, Z_2) &= \lambda_1 Z_1 G^1(Z_1, Z_2) + \left(\frac{\mu}{Z_1} G^1(Z_1, Z_2) - \mu G_1^1(Z_2)\right) \\ + \left(\frac{\mu}{Z_2} G^2(Z_1, Z_2) - \frac{\mu}{Z_2} G_0^2(Z_2)\right) &+ \lambda_1 Z_1 P_{0,0,0} \end{aligned}$$

$$(\lambda_1(1 - Z_1) + \lambda_2(1 - Z_2) + \mu(1 - \frac{1}{Z_1}))G^1(Z_1, Z_2) = \frac{\mu}{Z_2} G^2(Z_1, Z_2) + \lambda_1 Z_1 P_{0,0,0} - \mu G_1^1(Z_2) - \frac{\mu}{Z_2} G_0^2(Z_2)$$

De l'équation (1*), on tire la valeur de $\mu G_1^1(Z_2)$, et on trouve l'égalité suivante :

$$\mu G_1^1(Z_2) = (\lambda_1 + \lambda_2(1 - Z_2) + \mu(1 - \frac{1}{Z_1})) G_0^2(Z_2) + (\lambda_1 + \lambda_2(1 - Z_2)) P_{0,0,0}$$

Et de l'équation {(2*) et (3*)}, on tire la valeur $G^2(Z_1, Z_2)$, et on trouve la fraction suivante :

$$G^2(Z_1, Z_2) = \frac{(\lambda_1 + \lambda_2(1 - Z_2) + \mu)G_0^2(Z_2)}{\lambda_1(1 - Z_1) + \lambda_2(1 - Z_2) + \mu} \tag{a}$$

Maintenant on remplace $\mu G_1^1(Z_2)$ et $G^2(Z_1, Z_2)$ par leur valeur dans l'équation {(4*) + (5*)},

Pour obtenir $G^1(Z_1, Z_2)$

$$\begin{aligned}
 & (\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu(1-\frac{1}{Z_1}))G^1(Z_1, Z_2) = \frac{\mu}{Z_2} \left[\frac{(\lambda_1 + \lambda_2(1-Z_2) + \mu)G_0^2(Z_2)}{\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu} \right] + \lambda_1 Z_1 P_{0,0,0} \\
 & - \left[(\lambda_1 + \lambda_2(1-Z_2) + \mu(1-\frac{1}{Z_2}))G_0^2(Z_2) + (\lambda_1 + \lambda_2(1-Z_2))P_{0,0,0} \right] - \frac{\mu}{Z_2} G_0^2(Z_2) \\
 & = \left[\frac{\mu}{Z_2} * \frac{(\lambda_1 + \lambda_2(1-Z_2) + \mu)}{\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu} - (\lambda_1 + \lambda_2(1-Z_2) + \mu) \right] G_0^2(Z_2) - (\lambda_1(1-Z_1) + \\
 & \lambda_2(1-Z_2))P_{0,0,0}
 \end{aligned}$$

Alors on obtient :

$$G^1(Z_1, Z_2) = \frac{(\lambda_1 + \lambda_2(1-Z_2) + \mu)G_0^2(Z_2) \left[\frac{\mu}{Z_2} * \frac{1}{\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu} - 1 \right] - (\lambda_1(1-Z_1) + \lambda_2(1-Z_2))P_{0,0,0}}{\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu \left(1 - \frac{1}{Z_1} \right)} \quad (b)$$

En posant que le dénominateur est égal à 0 sous la condition que $|Z_1| = 1$ et $|Z_2| < 1$, on obtient la relation suivante :

$$\lambda_1(1-Z_1) + \lambda_2(1-Z_2) + \mu - \frac{\mu}{Z_1} = 0 \quad (c)$$

$$\Rightarrow \lambda_1 Z_1^2 - [\lambda_1 + \lambda_2(1-Z_2) + \mu] Z_1 + \mu = 0$$

Sachant que

$$|\lambda_1 Z_1^2 + \mu| \leq \lambda_1 + \mu < |\lambda_1 + \lambda_2(1-Z_2) + \mu|$$

Et comme :

$$|\lambda_1 Z_1^2 + \mu| \leq |\lambda_1 + \lambda_2(1-Z_2) + \mu|$$

Alors on conclut que l'équation (c) admet une solution unique qui est une fonction de Z_2 et ceci d'après le théorème de Rouché.

Pour que $G^1(Z_1, Z_2)$ converge il faut que $|Z_1| < 1$ et $|Z_2| < 1$, si l'on pose également le numérateur de $G^1(Z_1, Z_2)$ égal à 0, il sera donc évident que la solution de cette équation sera également une fonction de Z_2 ($g(Z_2)$).

$$[\lambda_1 + \lambda_2(1 - Z_2) + \mu]G_0^2(Z_2) \left[\frac{\mu}{Z_2} * \frac{1}{\lambda_1(1-Z_1)+\lambda_2(1-Z_2)+\mu} - 1 \right] - [\lambda_1(1 - Z_1) + \lambda_2(1 - Z_2)]P_{0,0,0} = 0$$

Si : $Z_1 = g(Z_2)$, alors

$$[\lambda_1 + \lambda_2(1 - Z_2) + \mu]G_0^2(Z_2) \left[\frac{\mu}{Z_2} \frac{1}{\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2)+\mu} - 1 \right] [\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2)]P_{0,0,0} = 0$$

$$\Rightarrow G_0^2(Z_2) = \frac{[\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2)+\mu][(\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2))]}{[\lambda_1+\lambda_2(1-Z_2)+\mu] \left[\frac{\mu}{Z_2} - (\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2)+\mu) \right]} P_{0,0,0}$$

$$G_0^2(Z_2) = \frac{[\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2)+\mu][(\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2))]}{[\lambda_1+\lambda_2(1-Z_2)+\mu] \left[\frac{\mu}{Z_2} - (\lambda_1(1-g(Z_2))+\lambda_2(1-Z_2)+\mu) \right]} P_{0,0,0} \quad (d)$$

En posant $Z_1 = g(Z_2)$ dans (c), on obtient :

$$\lambda_1 g^2(Z_2) - [\lambda_1 + \lambda_2(1 - Z_2) + \mu] g(Z_2) + \mu = 0$$

Et Si: $Z_2 = 1$, l'équation devient :

$$\lambda_1 g^2(1) - [\lambda_1 + \mu] g(1) + \mu = 0$$

On calcule le Δ pour trouver les solutions de cette équation :

$$\Delta = (\lambda_1 + \mu)^2 - 4\lambda_1 \mu = (\lambda_1 - \mu)^2 > 0$$

Alors :

$$g(1) = \frac{\lambda_1 + \mu + \lambda_1 - \mu}{2\lambda_1} = \frac{2\lambda_1}{2\lambda_1} = 1.$$

$$g(1) = \frac{\lambda_1 + \mu - \lambda_1 + \mu}{2\lambda_1} = \frac{2\mu}{2\lambda_1} = \frac{\mu}{\lambda_1} > 1 \text{ (à rejeter)}$$

Ainsi pour :

$$\lambda_1 g^2(Z_2) - [\lambda_1 + \lambda_2(1 - Z_2) + \mu] g(Z_2) + \mu = 0$$

On dérive une première fois cette équation au point $Z_2 = 1$, on trouve :

$$2\lambda_1 g'(1) g(1) - (-\lambda_2 g(1) + (\lambda_1 + \underbrace{\lambda_2(1-1)}_{=0} + \mu) g'(1)) = 0$$

$$= 2\lambda_1 g'(1)g(1) + \lambda_2 g(1) - (\lambda_1 + \mu) g'(1) = 0$$

On sait que : $g(1) = 1$, donc on remplace dans la formule précédente et on obtient :

$$(\lambda_1 - \mu) g'(1) + \lambda_2 = 0 \quad \Rightarrow \quad \boxed{g'(1) = \frac{\lambda_2}{(\mu - \lambda_1)}} \quad (e)$$

On procède également à la seconde dérivée de cette même équation et toujours au point $Z_2 = 1$, tout en remplaçant $g(1)$ et $g'(1)$ par leur valeur. Du coup on aura :

$$\boxed{g''(1) = \frac{2\lambda_2^2 \mu}{(\mu - \lambda_1)^3}} \quad (f)$$

On constate que lorsque Z_2 converge vers 1, le numérateur ainsi que le dénominateur de $G_0^2(Z_2)$ sont égaux à 0. De ce fait on applique la règle de l'hôpital à $G_0^2(Z_2)$ au point $Z_2 = 1$ tout en utilisant certaines astuces de calcul.

On a :

$$G_0^2(1) = \lim_{z_2=1} \frac{[\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2) + \mu][\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2)]}{[\lambda_1 + \lambda_2(1 - Z_2) + \mu] \left[\frac{\mu}{Z_2} - (\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2) + \mu) \right]} P_{0,0,0}$$

$$= \frac{\mu}{\lambda_1 + \mu} \lim_{z_2=1} \frac{[\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2)]}{\left[\frac{\mu}{Z_2} - (\lambda_1(1 - g(Z_2)) + \lambda_2(1 - Z_2) + \mu) \right]} P_{0,0,0}$$

Ainsi après dérivation du numérateur et du dénominateur on obtient :

$$G_0^2(1) = \frac{\mu}{\lambda_1 + \mu} * \frac{-\lambda_1 g'(Z_2) - \lambda_2}{\frac{-\mu}{Z_2^2} + \lambda_1 g'(Z_2) + \lambda_2} P_{0,0,0} \quad \text{au point } Z_2 = 1$$

En remplaçant Z_2 et $g'(Z_2)$ par leur valeur on aura :

$$G_0^2(1) = \frac{\mu\lambda_2}{(\lambda_1 + \mu)(\mu - \lambda_1 - \lambda_2)} P_{0,0,0} \quad (g)$$

Nous obtiendrons donc $G^1(1,1)$ et $G^2(1,1)$ en utilisant l'équation (g) et la règle de l'hôpital sur les équations (a) et (b).

$$\begin{aligned} G^2(1,1) &= \frac{\lambda_1 + \mu}{\mu} G_0^2(1) \\ &= \frac{\lambda_1 + \mu}{\mu} \left(\frac{\mu\lambda_2}{(\lambda_1 + \mu)(\mu - \lambda_1 - \lambda_2)} P_{0,0,0} \right) \end{aligned}$$

D'où

$$G^2(1,1) = \frac{\lambda_2}{\mu - \lambda_1 - \lambda_2} P_{0,0,0} \quad (h)$$

De la même manière on obtient $G^1(1,1)$

$$G^1(1,1) = \left(\frac{\mu - \lambda_2}{\mu - \lambda_1 - \lambda_2} - 1 \right) P_{0,0,0} \quad (i)$$

Sachant que

$$G^1(1,1) + G^2(1,1) + P_{0,0,0} = 1$$

Par conséquent nous aurons après calcul:

$$P_{0,0,0} = \frac{\mu - \lambda_1 - \lambda_2}{\mu} \quad (j)$$

Par la suite nous pouvons également calculer la première dérivée de $G_0^2(z_2)$ au point $z_2 = 1$, à partir de la formule de $G_0^2(z_2)$ et en remplaçant $g(1)$, $g'(1)$ et $g''(1)$ par leur valeur. On aura donc :

$$G_0^{2'}(1) = \frac{\mu\lambda_2[\lambda_1^3 + \lambda_1^2(3\lambda_2 - \mu) + \mu^3 + \lambda_1(2\lambda_2^2\mu - \mu\lambda_2 - \mu^2)]}{(\mu - \lambda_1 - \lambda_2)(\lambda_1 + \mu)^2(\mu - \lambda_1)} P_{0,0,0} \quad (k)$$

Les mesures de performance :

Le nombre moyen de clients dans le système :

Proposition : Le nombre moyen de clients prioritaire L_1 et celui du client non prioritaire L_2 dans le système sont données par :

$$\text{➤ } L_1 = \frac{\lambda_1(\lambda_2 + \mu)}{\mu(\mu - \lambda_1)} \quad (l)$$

$$\text{➤ } L_2 = \frac{\lambda_2(\mu^2 - \mu\lambda_1 + \lambda_1^2 + \lambda_1\lambda_2)}{\mu(\mu - \lambda_2 - \lambda_1)(\mu - \lambda_1)} \quad (m)$$

Preuve :

$$\bullet \quad L_1 = \left. \frac{\partial(G^1(z_1,1) + G^2(z_1,1))}{\partial z_1} \right|_{z_1=1}$$

Où

$$\left. \frac{\partial G^1(z_1,1)}{\partial z_1} \right|_{z_1=1} = \frac{\lambda_1^2\lambda_2 + \mu^2\lambda_1}{\mu^2(\mu - \lambda_1)} \quad ; \quad \left. \frac{\partial G^2(z_1,1)}{\partial z_1} \right|_{z_1=1} = \frac{\lambda_1\lambda_2}{\mu^2}$$

Par remplacement, on trouve :

$$L_1 = \frac{\lambda_1^2 \lambda_2 + \mu^2 \lambda_1}{\mu^2 (\mu - \lambda_1)} + \frac{\lambda_1 \lambda_2}{\mu^2} = \frac{\lambda_1 (\lambda_2 + \mu)}{\mu (\mu - \lambda_1)} ;$$

$$\bullet \quad L_2 = \left. \frac{\partial(G^1(1, z_2) + G^2(1, z_2))}{\partial z_2} \right|_{z_2=1}$$

Où

$$\left. \frac{\partial G^2(1, z_2)}{\partial z_2} \right|_{z_2=1} = \left[\left(\frac{\lambda_2}{\mu} - 1 \right) + \frac{\mu + \lambda_1}{\lambda_2} \left(1 + \frac{\lambda_2^2}{\mu^2} - \frac{\lambda_2}{\mu} \right) \right] G_0^2(1) + \left(1 - \frac{\lambda_2}{\mu} \right) \left(\frac{\mu + \lambda_1}{\lambda_2} \right) G_0^{2'}(1);$$

$$\left. \frac{\partial G^1(1, z_2)}{\partial z_2} \right|_{z_2=1} = \frac{\lambda_1 (\lambda_1 + \lambda_2 + \mu)}{\mu^2} G_0^2(1) + \frac{\mu + \lambda_1}{\mu} G_0^{2'}(1);$$

Sachant que $G_0^2(1)$ et $G_0^{2'}(1)$ sont respectivement données par les expressions (g) et (k).

Alors

$$\begin{aligned} L_2 &= \frac{\lambda_1^2 \lambda_2 + \mu^3 + \lambda_1 (2\lambda_2^2 + \mu^2)}{\mu^2 (\mu + \lambda_1)} + \frac{\mu (\mu^3 + \lambda_1^3 + \lambda_1^2 (3\lambda_2 - \mu) + \lambda_1 (2\lambda_2^2 \mu - \mu \lambda_2 - \mu^2))}{(\mu^2 - \lambda_1^2) (\mu - \lambda_1 - \lambda_2)^2} \\ &= \frac{\lambda_2 (\mu^2 - \mu \lambda_1 + \lambda_1^2 + \lambda_1 \lambda_2)}{\mu (\mu - \lambda_2 - \lambda_1) (\mu - \lambda_1)} \end{aligned}$$

Donc on en conclut que le nombre moyen de clients L dans le système est :

$$L = L_1 + L_2$$

Ainsi sous la condition de stationnarité du système ($\frac{\lambda_1 + \lambda_2}{\mu} < 1$), nous déduirons les autres mesures de performance et ceci par la formule de Little.

■ Le nombre moyen de clients L_q dans la file :

$$L_{q1} = \frac{\lambda_1 (\lambda_1 + \lambda_2)}{\mu (\mu - \lambda_1)} ; \quad L_{q2} = \frac{\lambda_2 (\lambda_1 + \lambda_2)}{(\mu - \lambda_2 - \lambda_1) (\mu - \lambda_1)}$$

- Le temps moyen d'attente w_q dans la file :

$$W_{q_1} = \frac{\lambda_1 + \lambda_2}{\mu(\mu - \lambda_1)} ; \quad W_{q_2} = \frac{\lambda_1 + \lambda_2}{(\mu - \lambda_2 - \lambda_1)(\mu - \lambda_1)} ;$$

- Le temps moyen de séjour W dans le système :

$$W_1 = \frac{\lambda_2 + \mu}{\mu(\mu - \lambda_1)} ; \quad W_2 = \frac{\mu^2 - \mu\lambda_1 + \lambda_1^2 + \lambda_1\lambda_2}{\mu(\mu - \lambda_2 - \lambda_1)(\mu - \lambda_1)} ;$$

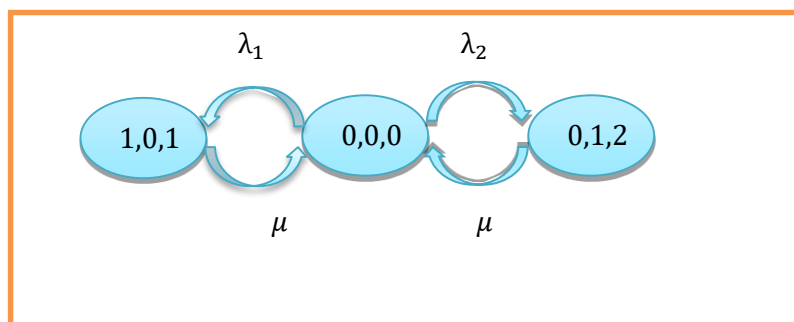
3.1.2 Système d'attente M/M/1 avec Priorité absolue

- **Description**

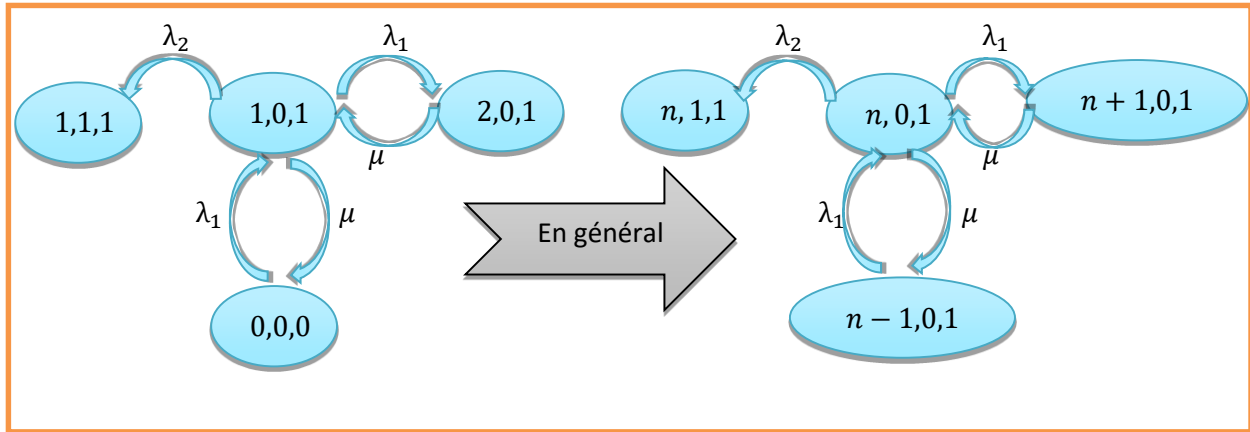
Pour la priorité absolue, un client prioritaire qui arrive et trouve l'unique serveur occupé par le service d'un autre client prioritaire, prend une place dans la file d'attente, contrairement au cas où il le trouve occupé par le service d'un client non prioritaire : il commence immédiatement son service, et ce client dont le service a été interrompu par l'arrivée du client prioritaire rejoint la file tout en étant à la tête afin d'être de nouveau pris en charge par le serveur à partir du point où il a été interrompu.

- **Les graphes de transition et les équations de Kolmogorov pour la priorité absolue**

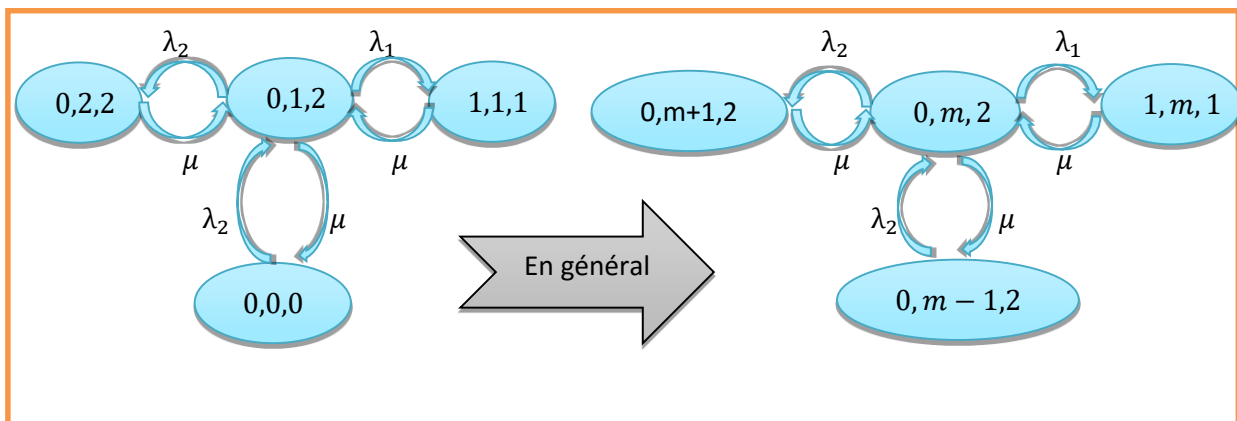
L'état (0, 0, 0) : représente l'état où le système est vide. Ainsi le graphe correspondant à cet état est donné par :



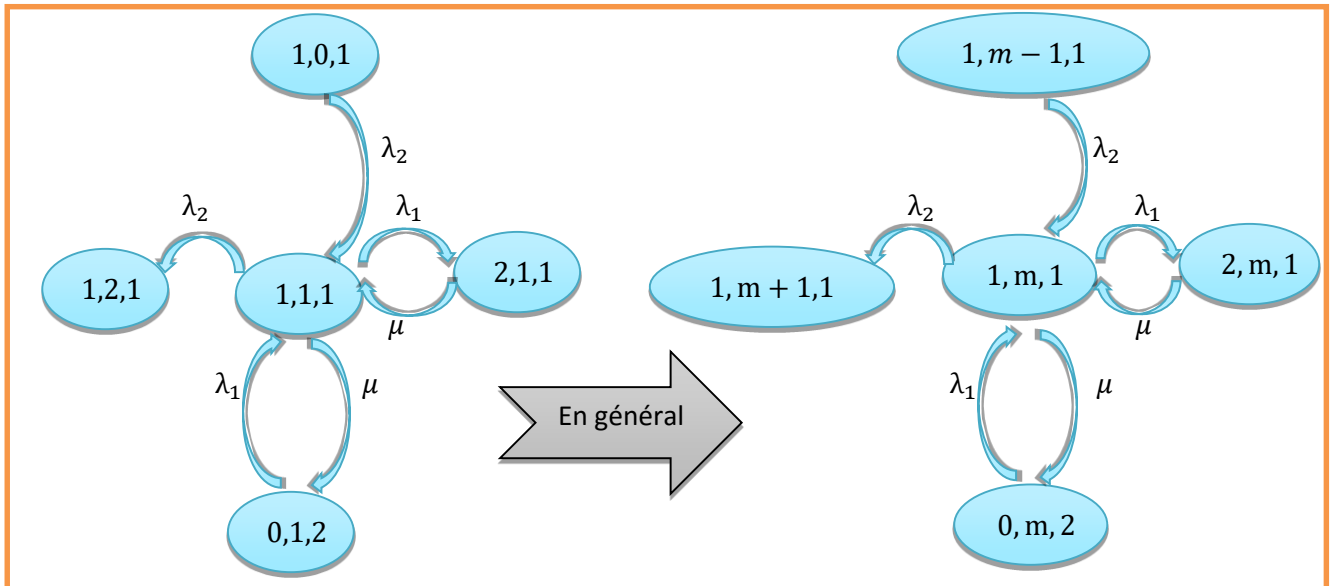
L'état $(1, 0, 1)$ et $(n, 0, 1)$: ces deux états représentent les cas où le système ne comporte uniquement qu'un (resp n) client prioritaire.



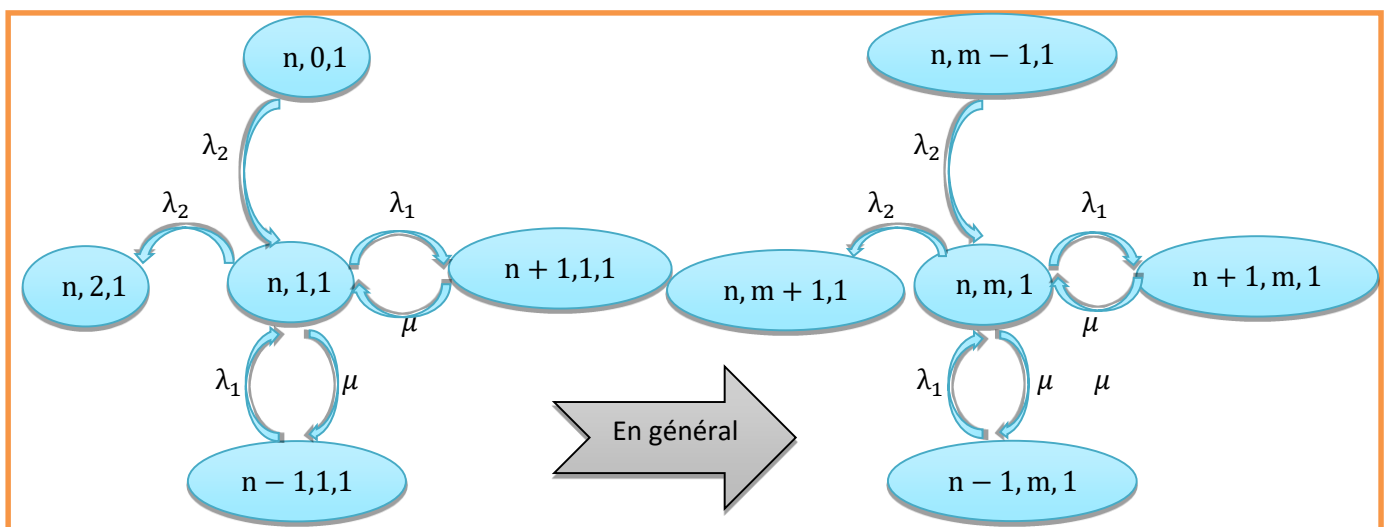
L'état $(0, 1, 2)$ et $(0, m, 2)$: ces deux états caractérisent les cas où le système contient 0 client prioritaire, un client (resp m) client non prioritaire et le serveur étant occupé par un client non prioritaire.



L'état (1, 1, 1) et (1, m, 1): ces graphes ci-dessous représentent les états (1, 1, 1) et (1, m, 1) qui caractérisent le système lorsqu'il contient un client prioritaire, et un (resp m) client non prioritaire, avec un client prioritaire au serveur.



L'état (n, 1, 1) et (n, m, 1) : ces états décrivent le cas où le système contient un (resp m) client non prioritaire et que le serveur soit occupé par le service de l'un des n clients prioritaires présents dans le système. On les représente graphiquement de la manière suivante :



A L'aide des graphes de transition précédents nous obtenons le système d'équation(les équations de Kolmogorov) suivant :

$$(\lambda_1 + \lambda_2)P_{0,0,0} = \mu P_{1,0,1} + \mu P_{0,1,2} \quad (1)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{0,1,2} = \lambda_2 P_{0,0,0} + \mu P_{0,2,2} + \mu P_{1,1,1} \quad (2)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{0,m,2} = \lambda_2 P_{0,m-1,2} + \mu P_{0,m+1,2} + \mu P_{1,m,1} \quad (3)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,0,1} = \lambda_1 P_{0,0,0} + \mu P_{2,0,1} \quad (4)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,1,1} = \lambda_2 P_{1,0,1} + \mu P_{2,1,1} + \lambda_1 P_{0,1,2} \quad (5)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{1,m,1} = \lambda_2 P_{1,m-1,1} + \mu P_{2,m,1} + \lambda_1 P_{0,m,2} \quad (6)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,0,1} = \lambda_1 P_{n-1,0,1} + \mu P_{n+1,0,1} \quad (7)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,1,1} = \lambda_2 P_{n,0,1} + \lambda_1 P_{n-1,1,1} + \mu P_{n+1,1,1} \quad (8)$$

$$(\lambda_1 + \lambda_2 + \mu)P_{n,m,1} = \lambda_2 P_{n,m-1,1} + \lambda_1 P_{n-1,m,1} + \mu P_{n+1,m,1} \quad (9)$$

▪ **La résolution :**

Par l'addition des équations on trouve :

(1) +(2) +(3) donne :

$$(\lambda_1 + \lambda_2 + \mu) \sum_{m=1}^{\infty} Z_2^m P_{0,m,2} = \lambda_2 Z_2 G_0^2(Z_2) + \mu G_1^1(Z_2) + \frac{\mu}{Z_2} G_0^2(Z_2)$$

$$(\lambda_1 + \lambda_2(1 - Z_2) + \mu(1 - \frac{1}{Z_2}))G_0^2(Z_2) = \mu G_1^1(Z_2) - (\lambda_1 + \lambda_2(1 - Z_2))P_{0,0,0}$$

(a)

(4)+(5) +(6) donne :

$$(\lambda_1 + \lambda_2 + \mu)G_1^1 (Z_2) = \lambda_1 P_{0,0,0} + \lambda_1 G_0^2 (Z_2) + \lambda_2 Z_2 G_1^1 (Z_2) + \mu G_2^1 (Z_2)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu)G_1^1 (Z_2) = \lambda_1 P_{0,0,0} + \lambda_1 G_0^2 (Z_2) + \mu G_2^1 (Z_2) \quad (b)$$

(7)+(8) +(9) donne :

$$(\lambda_1 + \lambda_2 + \mu)G_n^1 (Z_2) = \lambda_1 G_{n-1}^1 (Z_2) + \lambda_2 Z_2 G_n^1 (Z_2) + \mu G_{n+1}^1 (Z_2)$$

$$(\lambda_1 + \lambda_2(1- Z_2) + \mu)G_n^1 (Z_2) = \lambda_1 G_{n-1}^1 (Z_2) + \mu G_{n+1}^1 (Z_2) \quad (c)$$

(b)+ (c) donne :

$$(\lambda_1 + \lambda_2 (1- Z_2) + \mu) G^1 (Z_1, Z_2) = \lambda_1 Z_1 P_{0,0,0} + \lambda_1 Z_1 G^1 (Z_1, Z_2) + \frac{\mu}{Z_1} G^1 (Z_1, Z_2) - \mu G_1^1 (Z_2) + \lambda_1 Z_1 G_0^2 (Z_2)$$

En remplaçant $\mu G_1^1 (Z_2)$ par sa valeur tirée dans (a)

$$\Rightarrow (\lambda_1(1 - Z_1) + \lambda_2(1- Z_2) + \mu(1- \frac{1}{Z_1}))G^1 (Z_1, Z_2) = \lambda_1 Z_1 P_{0,0,0} + \lambda_1 Z_1 G_0^2 (Z_2)$$

$$- (\lambda_1 + \lambda_2(1- Z_2))P_{0,0,0} - (\lambda_1 + \lambda_2(1- Z_2) + \mu(1- \frac{1}{Z_2}))G_0^2 (Z_2)$$

$$\Rightarrow (\lambda_1(1 - Z_1) + \lambda_2(1- Z_2) + \mu(1- \frac{1}{Z_1}))G^1 (Z_1, Z_2) = - (\lambda_1(1- Z_1) + \lambda_2(1- Z_2)$$

$$+ \mu(1- \frac{1}{Z_2}))G_0^2 (Z_2) - (\lambda_1(1- Z_1) + \lambda_2(1- Z_2))P_{0,0,0}$$

$$G^1(Z_1, Z_2) = \frac{-(\lambda_1(1-Z_1)+\lambda_2(1-Z_2)+\mu(1-\frac{1}{Z_2}))G_0^2(Z_2)-(\lambda_1(1-Z_1)+\lambda_2(1-Z_2))P_{0,0,0}}{(\lambda_1(1-Z_1)+\lambda_2(1-Z_2)+\mu(1-\frac{1}{Z_1}))} \quad (d)$$

Et en posant le numérateur de (d) égal à zéro on aura :

$$G_0^2(Z_2) = \frac{-(\lambda_1(1-Z_1)+\lambda_2(1-Z_2))P_{0,0,0}}{(\lambda_1(1-Z_1)+\lambda_2(1-Z_2)+\mu(1-\frac{1}{Z_2}))} \quad (e)$$

En utilisant la règle de l'hôpital sur (e) on obtient :

$$G_0^2(Z_2) = \lim_{Z_2 \rightarrow 1} \frac{(\lambda_1 g'(Z_2) + \lambda_2)P_{0,0,0}}{-\lambda_1 g'(Z_2) - \lambda_2 + \frac{\mu}{Z_2^2}} = \frac{\mu \lambda_2 P_{0,0,0}}{\mu^2 - \mu \lambda_2 - \mu \lambda_1}$$

On aura :

$$G_0^2(1) = \frac{\lambda_2}{\mu - \lambda_1 - \lambda_2} P_{0,0,0} \quad (f)$$

Nous obtiendrons donc $G^1(1,1)$ en utilisant (f) et la règle de l'hôpital sur (d)

$$G^1(1,1) = \left(\frac{\mu}{\mu - \lambda_1 - \lambda_2} - 1\right) P_{0,0,0} \quad (g)$$

Ainsi sachant que :

$$G^1(1,1) + P_{0,0,0} = 1$$

Après calcul on obtient :

$$P_{0,0,0} = \frac{\mu - \lambda_1 - \lambda_2}{\mu} \quad (h)$$

▪ Mesures de performance

Le système complet (avec clients prioritaires et non prioritaires) étant considéré comme un système d'attente M/M/1 avec $(\rho = \frac{\lambda_1 + \lambda_2}{\mu} < 1)$ comme taux de stabilité, on peut donc écrire d'après le chapitre 1 :

$$L = \frac{\rho}{1-\rho} \quad \text{avec } \rho = 1 - P_{0,0,0} \quad (i)$$

Où L représente le nombre moyen de client (prioritaire et non prioritaire) dans le système.

En observant également que les clients non prioritaires n'affectent en aucun cas la file des clients prioritaires alors on peut donc conclure que les clients prioritaires constituent dans ce cas un système de file d'attente classique M/M/1 avec $(\rho_1 = \frac{\lambda_1}{\mu} < 1)$ comme taux de stabilité.

Soit P_1 la probabilité que le serveur soit occupé par le service d'un client prioritaire

$$P_1 = \rho_1$$

Dans ce cas d'après le chapitre 1 :

➤ Le nombre moyen de clients prioritaire dans le système L_1 est

$$L_1 = \frac{\rho_1}{1-\rho_1} \quad \text{avec } \rho_1 = \frac{\lambda_1}{\mu} \quad (j)$$

➤ Le nombre moyen de clients non prioritaire dans le système L_2 peut être obtenue en soustrayant le nombre moyen de clients prioritaire dans le système L_1 du nombre moyen de clients dans le système L.

$$L_2 = L - L_1 = \left(\frac{\rho_1 + \rho_2}{1 - \rho_1 - \rho_2} \right) - \left(\frac{\rho_1}{1 - \rho_1} \right)$$

D'où

$$L_2 = \frac{\rho_2}{(1-\rho_1-\rho_2)(1-\rho_1)} \quad \text{avec} \quad \rho_2 = \frac{\lambda_2}{\mu} < 1 \quad \text{et} \quad \rho_1 = \frac{\lambda_1}{\mu} \quad (k)$$

Au moyen des formules de Little et par déduction de (j) et de (k) on peut obtenir les autres mesures de performances.

❖ Le nombre moyen de client Lq dans la file :

$$Lq_1 = \frac{\lambda_1^2}{\mu(\mu-\lambda_1)} \quad ; \quad Lq_2 = \frac{\lambda_2(2\lambda_1\mu+\lambda_2\mu-\lambda_1^2-\lambda_1\lambda_2)}{\mu(\mu-\lambda_2-\lambda_1)(\mu-\lambda_1)}$$

❖ Le temps moyen d'attente Wq dans la file :

$$Wq_1 = \frac{\lambda_1}{\mu(\mu-\lambda_1)} \quad ; \quad Wq_2 = \frac{2\lambda_1\mu+\lambda_2\mu-\lambda_1^2-\lambda_1\lambda_2}{\mu(\mu-\lambda_2-\lambda_1)(\mu-\lambda_1)}$$

❖ Le temps moyen de séjour W dans le système :

$$W_1 = \frac{1}{(\mu-\lambda_1)} \quad ; \quad W_2 = \frac{\mu}{(\mu-\lambda_2-\lambda_1)(\mu-\lambda_1)} ;$$

3.2 Les illustrations numériques :

Dans cette section, nous illustrons numériquement l'influence des taux d'arrivées des clients (λ_1 et λ_2) sur certaines mesures de performance des deux systèmes étudiés dans la première partie de ce chapitre dans le but de faire une comparaison entre ces derniers. Afin d'obtenir ces résultats nous avons donc utilisé les logiciels Matlab et Excel.

3.2.1 L'effet de la variation du taux λ_1 sur le temps moyen d'attente des clients dans la file pour les deux systèmes (M/M1 avec priorité relative et M/M/1 avec priorité absolue):

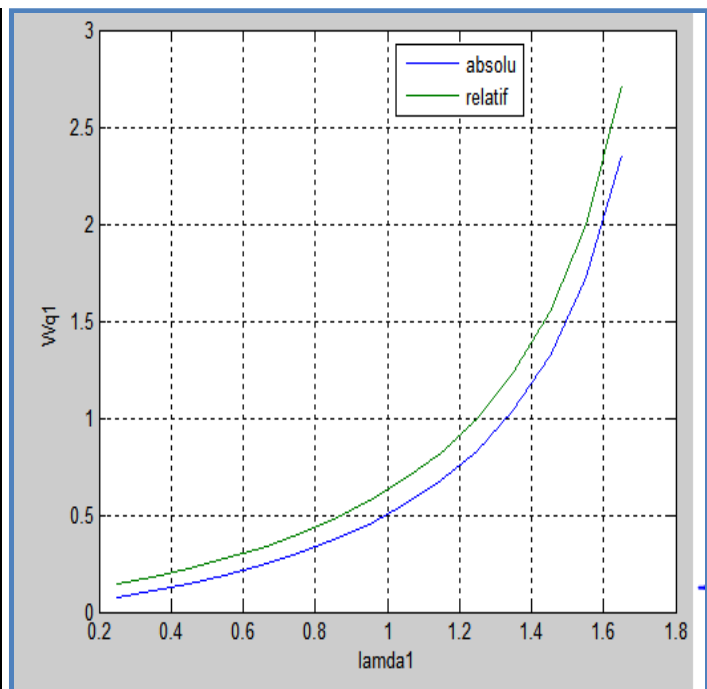
On choisit les valeurs arbitraires suivantes :

$\lambda_2 = 0,25$; **et** $\mu = 2$;

On fait varier le taux des arrivées prioritaires λ_1 pour les différentes durées d'attente dans la file tout en respectant la condition de stationnarité c'est-à-dire ($\rho = \frac{\lambda_1 + \lambda_2}{\mu} < 1$) .

◆ Pour les clients prioritaires :

λ_1	W_{q1} Absolu	W_{q1} Relatif
0,25	0,07142857	0,14285714
0,35	0,10606061	0,18181818
0,45	0,14516129	0,22580645
0,55	0,18965517	0,27586207
0,65	0,24074074	0,33333333
0,75	0,3	0,4
0,85	0,36956522	0,47826087
0,95	0,45238095	0,57142857
1,05	0,55263158	0,68421053
1,15	0,67647059	0,82352941
1,25	0,83333333	1
1,35	1,03846154	1,23076923
1,45	1,31818182	1,54545455
1,55	1,72222222	2
1,65	2,35714286	2,71428571

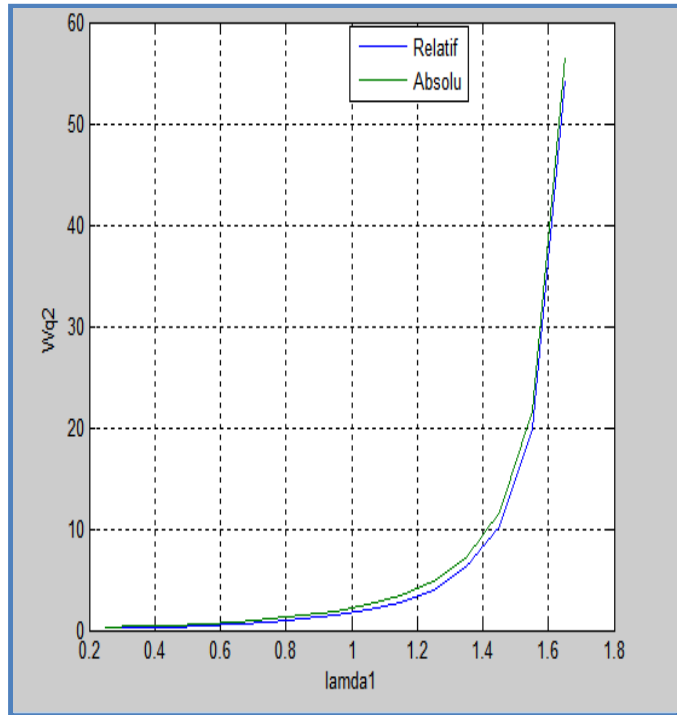


■ **Tableau 09** : l'effet de λ_1 sur W_{q1} (absolu et relative).

■ **Figure 15** : Variation de W_{q1} (absolu et relative) en fonction de λ_1

◆ Pour les clients non prioritaires :

λ_1	W_{q_2} Absolu	W_{q_2} Relatif
0,25	0,26190476	0,19047619
0,35	0,36580087	0,25974026
0,45	0,49255583	0,34739454
0,55	0,64942529	0,45977011
0,65	0,84680135	0,60606061
0,75	1,1	0,8
0,85	1,43236715	1,06280193
0,95	1,88095238	1,42857143
1,05	2,5075188	1,95488722
1,15	3,42156863	2,74509804
1,25	4,83333333	4
1,35	7,19230769	6,15384615
1,45	11,6212121	10,3030303
1,55	21,7222222	20
1,65	56,6428571	54,2857143



■ **Tableau 10** : l'effet de λ_1 sur W_{q_2} (absolu et relative).

■ **Figure 16** : Variation de W_{q_2} (absolu et relatif) en fonction de λ_1

3.2.2 L'effet de la variation du taux λ_2 sur le temps moyen d'attente des clients dans la file pour les deux systèmes (M/M1 avec priorité relative et M/M/1 avec priorité absolue):

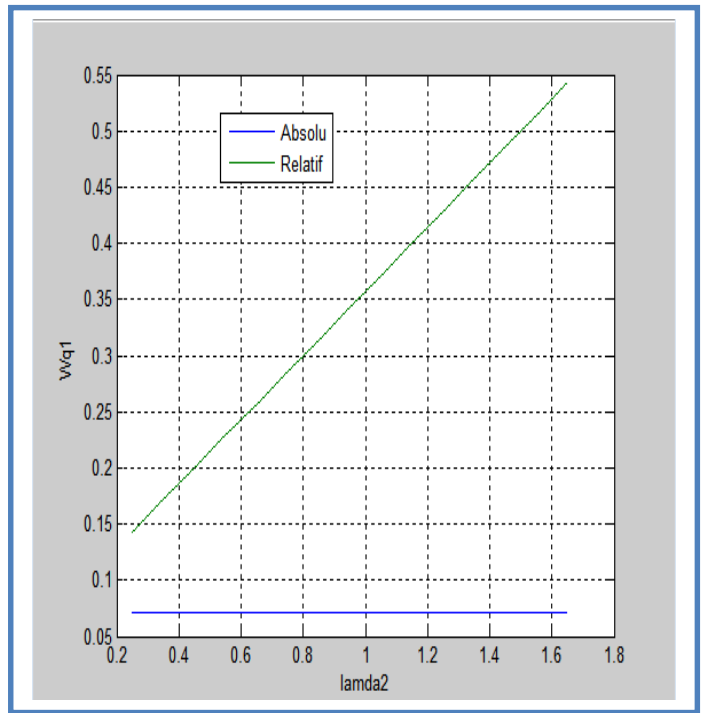
Cette fois-ci nous allons faire varier le taux des arrivées non prioritaires λ_2 tout en fixant le taux des arrivées prioritaires λ_1 et les taux de service μ .

On choisit alors

$$\lambda_1 = 0,25 \quad \text{et} \quad \mu = 2$$

◆ Pour les clients prioritaires :

λ_2	W_{q_1} Absolu	W_{q_1} Relatif
0,25	0,07142857	0,14285714
0,35	0,07142857	0,17142857
0,45	0,07142857	0,2
0,55	0,07142857	0,22857143
0,65	0,07142857	0,25714286
0,75	0,07142857	0,28571429
0,85	0,07142857	0,31428571
0,95	0,07142857	0,34285714
1,05	0,07142857	0,37142857
1,15	0,07142857	0,4
1,25	0,07142857	0,42857143
1,35	0,07142857	0,45714286
1,45	0,07142857	0,48571429
1,55	0,07142857	0,51428571
1,65	0,07142857	0,54285714

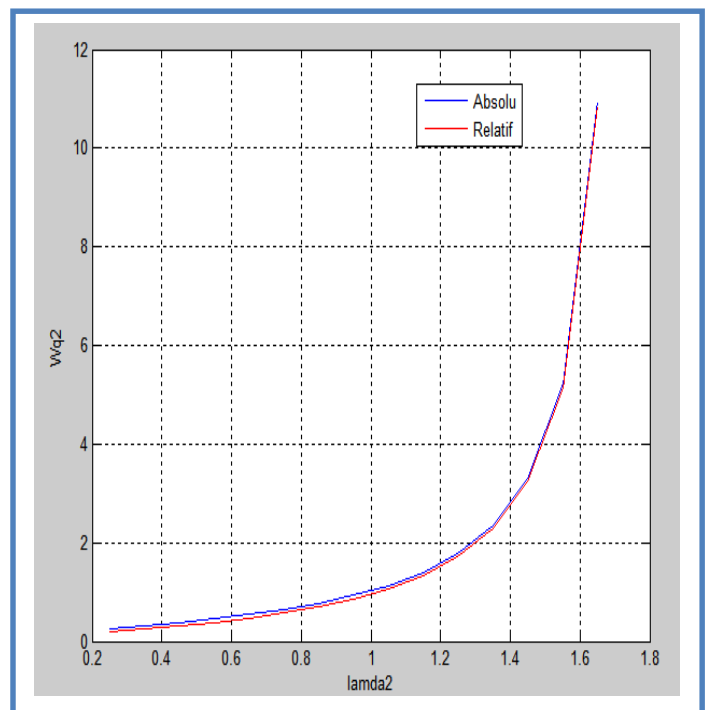


■ **Tableau 11:** l'effet de λ_2 sur W_{q_1} (absolu et relative).

✚ **Figure 17 :** Variation de W_{q_1} (absolu et relative) en fonction de λ_2

❖ Pour les clients non prioritaires :

λ_2	W_{q_2} Absolu	W_{q_2} Relatif
0,25	0,26190476	0,19047619
0,35	0,31632653	0,24489796
0,45	0,37912088	0,30769231
0,55	0,45238095	0,38095238
0,65	0,53896104	0,46753247
0,75	0,64285714	0,57142857
0,85	0,76984127	0,6984127
0,95	0,92857143	0,85714286
1,05	1,13265306	1,06122449
1,15	1,4047619	1,33333333
1,25	1,78571429	1,71428571
1,35	2,35714286	2,28571429
1,45	3,30952381	3,23809524
1,55	5,21428571	5,14285714
1,65	10,9285714	10,8571429



■ **Tableau 12:** l'effet de λ_2 sur W_{q_2} (absolu et relative).

✚ **Figure 18 :** Variation de W_{q_2} (absolu et relative) en fonction de λ_2 .

3.2.3 Interprétations

A partir des tableaux et figures ci-dessus et sous les conditions préalablement définies, on observe que :

- ◆ La variation du taux d'arrivée λ_1 influe sur le temps d'attente d'un client dans la file de la manière suivante :
 - un client jouissant d'une priorité absolue passe moins de temps dans la file d'attente par rapport à un client prioritaire dans un système d'attente avec priorité relative.
 - l'attente dans la file sera moins longue pour les clients non prioritaires des systèmes dont la politique de priorité est relative que pour les clients moins prioritaires des systèmes avec priorité absolue.
- ◆ La variation du taux d'arrivée λ_2 :
 - Elle n'influe en aucun cas sur le temps que passe un client jouissant d'une priorité absolue à attendre son service.
 - Les clients non prioritaires des deux systèmes passeront presque le même temps en attente avant d'être servis mais ceux des systèmes ayant une politique relative de priorité termineront donc leur service avec une légère durée d'avance.

CONCLUSION GENERALE ET PERSPECTIVES

Les systèmes de files d'attente occupent une place particulière dans la gestion et l'exploitation des systèmes réseaux permettant le bon fonctionnement des services publics

Dans ce travail nous nous sommes intéressés aux systèmes des files d'attente M/M/1 avec deux types de priorités (relative et absolue).

Dans un premier temps, nous avons commencé par la définition des différentes notions théoriques concernant les systèmes de file d'attente (processus de naissance et de mort, chaîne de Markov, notation du Kendall ...), puis comme une entrée sur notre sujet, nous avons effectué une étude bien détaillée sur quelques systèmes markoviens et semi-markoviens.

Ensuite dans un deuxième temps, nous nous sommes penchés sur la description des divers cas d'interruption de service, pour cela nous avons donc pris comme exemples : arrivées négative, serveur en vacance, serveur en panne etc.

Et finalement, dans la dernière partie de notre mémoire nous avons effectué une analyse mathématique du système de file d'attente M/M/1 avec deux types de priorités : priorité relative et absolue. A cet effet, nous avons réalisés les différents graphes de transition d'état des deux systèmes à partir desquels nous avons déduis les équations de Kolmogorov qu'on a ainsi utilisé par la suite pour déterminer les fonctions génératrices puis les mesures de performance.

Et pour confirmer nos résultats, on a fait une comparaison entre la priorité relative et absolue à l'aide de quelques illustrations numériques sur les différents temps d'attente.

Le travail que nous avons présenté dans ce mémoire nous a permis d'envisager plusieurs axes de recherches, notamment :

- Analyser les systèmes d'attente avec d'autres types de facteur d'interruption comme par exemples les systèmes avec les arrivées négatives ;
- Elargir le modèle présenté dans ce travail en associant un nouveau facteur d'interruption comme la panne du serveur par exemple.
- L'extension sur les modèles semi-Markoviens.
- Etudier les systèmes réseaux avec priorités.

Bibliographie

- [1] M. Houda and D. Taoufik, “étude de l’attente prioritaire dans les ports” Id : “ hal-00258630, 2009.
- [2] A.Cobham. Priority assignement in waiting line problems. *Operations Res*, 2 :70-76, 1954
- [3] F.F.Stephan. “ Two Queues inder pre-emptive priority with poisson arrival and service rates” *J.Opens.Res.Soc. Amer*, 4, 213-220.
- [4] G. Soraya, “Etude de files d’attente avec serveur non fiable, service groupé et vacances multiples,” Mémoire de magistère, USTHB, Alger 2014.
- [5] G.Rupert, Miller. Jr, “ Prioity Queues. *The Annals of Mathematical Statistics*“, Vol.31(N°1) :86-103, Mars 1960.
- [6] H.White and L.S. Christie, “ Queueing with preemptive or with breakdown.*Operations “*, *Res*, 6 :79-95, 1958.
- [7] J.Keilson, “ Queues subject to service interruption“, *Ann. Math Statis*, 33 :1314-1322, 1962.
- [8] J.L. Holley, “ Waiting line subject to priorities. *Operations*“, *Res*, 2 :341-343, 1954.
- [9] N.K.Jaiswal, “Prioity Queue“, *Academic Press New York and London, University of Southern Califournia*, 1968.
- [10] S.A. Dressin and E.Reich, “ Priority assignement on a waiting line“, *Quart Appel Math*, 15 :208-211, 1957.
- [11] K. R. Abdel-Mohcen, “Système de file d’attente markovien avec vacances simple,” Mémoire de master, université Saida, 2020.
- [12] Y. Yasmine et Z. Houda, “Evaluation des performances d’un mécanisme d’economie d’énergie d’un réseau de capteurs sans fil,” Mémoire de master, université Saad Dahleb de Blida, 2019.
- [13] J-F.Ducré-.Robitaille, “Files d’attente avec clients négatifs,” Mémoire de maitre en science, université Laval, 2000.
- [14] S. K. M. Muthukrishnan and S. R. Chakravarthy, “Preemptive Resume Priority Retrial Queue with Two Classes of MAP Arrivals,” doi: 10.12988/ams.2013.13231, January, 2013.
- [15] N.Oukid, Cours de file d’attente pour Master1. Université Saad Dahleb de Blida, 2021.
- [16] L. Bouterfa “Modèle de files avec rappels et priorité et leurs application,” thèse de doctorat, université Badji Mokhtar Annaba, 2015.

[17] H.Belkacemi et K.Bouzidi “Analyse et simulation de la communication RF dans les réseaux de capteurs via les files d’attente avec priorité,” Mémoire de master, Université de Bejaïa, 2017.

[18] G. Ayyappan and P. Thamizhselvi, “Priority Queueing System with a Single Server Serving Two Queues $M^{[X_1]}, M^{[X_2]}/G_1, G_2 /1$ with Balking and Optional Server Vacation,” Appl. Appl. Math, vol. 11, no. 1, pp. 61–82, 2016.