République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Hadj- Lakhdar -Batna

Faculté de Technologie

Département de Génie Industriel

MÉMOIRE DE MAGISTER

PRÉSENTÉE AU
Laboratoire d'Automatique et Productique
En vue de l'obtention du Diplôme de MAGISTER
Spécialité : Génie Industriel
Par

Kebbas Salah
Ingenieur en Genie Industriel

Thème

Contribution à la Correction et l'Amélioration de la Qualité de Service dans une Entreprise Publique, en utilisant les Réseaux de Files d'Attente

JYRY

Dr.Samir Abdelhamid	MC.A	Université Batna	Président
Dr.Djamel Mouss	MC.A	Université Batna	Rapporteur
Dr.Hayet Melakhessou	MC.A	Université Batna	Examinateur
Pr.Chaoui Allaoua	Pr	Université Constantine	Examinateur

Année 2013

Dédicaces

Je tien à dédier ce modeste travail aux êtres les plus chers à mon cœur :

À celle qui a veillé des nuits pour mon bien être, à celle que j'adore : Ma mère.

À mon père qui m'a guidé vers les voies de la réussite.

À mes frères et mes sœurs

À ma famille et mes amies

SALAH

Remerciements

Ce travail a été effectué au sein du département Génie Industriel de l'université de Batna .Tout d'abord, louange à dieu qui, grâce à son aide et sa volonté m' a amené au bout de ce travail en me préservant la vie saine pendant toutes mes études.

Je tiens à remercier très sincèrement toutes les personnes qui de près ou de loin m'ont aidé à réaliser ce travail. Ces remerciements sont adressés principalement à mon directeur de recherche, Monsieur Djamel Mouss pour sa grande disponibilité et son soutien moral et intellectuel au cours de la réalisation de ce projet.

Je suis honorée de la présence dans le jury de M.Samir Abdelhamid, Maitre de conférences à l'université Hadj Lakhdar Batna, de Mme . Hayet Melakhessou, Maitre de conférences à l'université Hadj Lakhdar Batna, de M. Chaoui Allaoua, Professeur à l'université Constantine. Qu'ils trouvent ici l'expression de mon profond respect.

Mes remerciements vont également à tous les membres de département Génie Industriel et encore à tous mes amis.

SALAH

TABLE DE MATIERES

Résume

Le travail présenté dans ce mémoire aborde le problème de l'amélioration de la qualité de service dans le secteur publique par l'utilisation des réseaux de files d'attente. La Théorie des files d'attente est une technique de la Recherche opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et de déterminer ses caractéristiques pour aider les gestionnaires dans leurs prises de décisions.

Généralement les files d'attentes se forment lorsque les clients arrivent de façon aléatoire pour se faire servir. L'exemple le plus courant de la vie de tous les jours est le problème d'attente devant d'un carrefour à feux. Ce dernier constitue un point sensible dans la circulation routière, il est touché d'une manière directe par la congestion du trafic. La commande automatique des carrefours à feux est initialement apparue pour pallier quelques problèmes de sécurité, caractéristiques du franchissement des intersections (gestion des points de conflit, files d'attente, …). Cependant, avec l'augmentation constante du volume de trafic, les objectifs de cette commande se sont étendus: on souhaite non seulement assurer le respect des contraintes de fonctionnement, mais aussi minimiser les filles d'attente dans le but de fluidifier au maximum le trafic. Généralement, les clients voient dans l'attente une activité sans valeur ajoutée et, s'ils attendent trop longtemps, ils associent cette perte de temps à une mauvaise qualité de service.

Dans ce contexte notre travail s'inscrit dans le cadre des méthodes utilisant les modèles discrets par nature. Il est divisé en quatre chapitres.

Nous avons commencé dans le premier chapitre par la présentation des notions de base concernant la qualité de service dans le secteur publique (Définition de la qualité, les enjeux de la qualité de service, le cycle de la qualité de service, le Mangement par la Qualité Totale,…etc.).

Ensuite, dans le deuxième chapitre nous avons mis l'accent sur les processus stochastiques qui servent comme outil mathématique de base pour résoudre les problèmes de files d'attente (le processus markovienne et processus de naissance et de mort qui sont des préliminaires nécessaires à l´etude des systèmes de files d'attente).

En effet Les processus stochastiques (ou aléatoires) permettent de modéliser des systèmes dont le comportement n'est que partiellement prévisible. La théorie est fondée sur le calcul des probabilités et les statistiques.

Puis, dans le troisième chapitre nous avons présenté les notions de base exprimant les phénomènes d'attente, nous avons présenté aussi les différents modèles et réseaux de files d'attente (RFA) qui ont une très grande importance en recherche opérationnelle. Ils servent à modéliser et d'analyser des systèmes physiques de type clients/serveurs. Ils permettent ainsi d'évaluer les performances et de mieux comprendre le comportement de ces systèmes. Nous avons ensuite défini les éléments importants à considérer dans la mesure de performance d'un système de file d'attente (nombre moyen de clients dans le système, nombre moyen de clients qui attendent d'être servis, temps moyen d'attente en file, …etc.).

Dans le dernier chapitre nous avons exploité la théorie des files d'attente pour l'évaluation de la performance d'un carrefour à feux. En effet, l'évaluation de la performance d'un système réel se décompose d'une étape de modélisation permettant de passer du système au modèle et d'une étape d'analyse des performances du modèle. Dans cette section, nous avons expliqué l'équation récurrente qui représente, d'une manière générale, la relation entre les longueurs des files d'attente durant une succession de cycles pendant la période de sursaturation et ensuite, nous avons représenté le système par un model discret grâce une nouvelle formulation sur les files d'attente. Le choix de ce modèle discret est dû à la cohérence entre les périodes dans lesquelles évoluent le système et les changements des signalisations. Enfin, nous avons exploité le principe du Maximum pour définir la stratège de la commande optimale du système. Enfin, nous avons présenté la simulation comme technique complémentaire pour les réseaux de files d'attente.

Introduction générale

Bien qu'elle ait déjà été largement utilisée dans les milieux industriels depuis la décennie précédente, la notion de qualité ne cesse d'être évoquée aujourd'hui par les services publics. En effet, les années 1980 ont vu, sur le plan mondial, s'intensifier un mouvement lancé depuis plusieurs années, qui tend à insuffler dans toutes les activités de production ou de service, des comportements et des méthodes propres à obtenir et garantir la qualité de manière régulière [1].

La qualité s'avère être un moyen d'assurer le développement et la survie des services publics et autres qui sont confrontés à des comparaisons avec des entreprises, dans la mesure où elle contribue à renouveler en profondeur les modes de légitimation et les principes d'organisation. Sans compter qu'une mauvaise qualité augmente les coûts de production des produits, mais aussi des services, y compris ceux des services publics, ce qui affecte directement les dépenses du gouvernement. Ainsi, l'efficacité des organisations publiques et donc, la qualité de leurs prestations, constitue un enjeu de taille dans la mesure où elle affecte la qualité de vie de tout les partenaires sociaux [1].

La qualité est de ce fait une condition indispensable pour présenter un produit ou service sur le marché. Pour obtenir cette qualité, l’entreprise doit structurer son organisation autour d’un modèle, souvent la norme ISO 9001. Mais elle doit également mettre en œuvre des méthodes et des outils pour accroître sa performance [1].

Pour évaluer la performance d’un système, on utilise soit les méthodes analytiques, telles que les réseaux de files d’attentes, soit la simulation. Chacune de ces méthodes comporte ses avantages et ses inconvénients. Les solutions analytiques bénéficient de temps de résolution très rapides. Les résultats peuvent être immédiats, car ils sont déterminés à partir d’équations mathématiques issues du formalisme emprunté [6].

L’utilisation des files d’attente remonte au début des années 1950 (Baynat 2000). Ce formalisme a été essentiellement utilisé pour la modélisation des systèmes industriels. Mais, de nos jours, il représente un outil de modélisation analytique de plusieurs types de systèmes et notamment les systèmes de service tels que dans l’optimisation de trafic. La théorie des files d’attente est une technique de la Recherche

Opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et de déterminer ses caractéristiques pour aider les gestionnaires dans leurs prises de décisions. (Philipe, 1999).

Les files d'attentes se forment lorsque les clients arrivent de façon aléatoire pour se faire servir. Les exemples les plus courants de la vie de tous les jours sont les caisses des supermarchés, les établissements de restauration rapide, les billetteries des aéroports, les cinémas, les bureaux de poste, les banques. Toutefois, lorsqu'on parle d'attente, on pense souvent à des personnes. Or, les « clients » en attente sont aussi des commandes en attente de traitement, des camions en attente de chargement ou de déchargement, des machines en attente de réparation, des programmes d'ordinateur qui attendent d'être exécutés, des avions qui attendent l'autorisation de décoller, des bateaux qui attendent les remorqueurs pour accoster, les voitures aux panneaux d'arrêt, les patients dans les salles d'urgence, etc.

Généralement, les clients voient dans l'attente une activité sans valeur ajoutée et, s'ils attendent trop longtemps, ils associent cette perte de temps à une mauvaise qualité de service. De la même façon, au sein de l'entreprise, des employés inoccupés ou des équipements inutilisés représentent des activités sans valeur ajoutée. Pour éviter ces situations, la majorité des entreprises ont mis en place des processus d'amélioration continue dont le but ultime est l'élimination de toute forme de gaspillage, notamment l'attente. Tous ces exemples révèlent l'importance de l'analyse des files d'attente.

Pour cela notre mémoire va être structuré autour de quatre chapitres organisés comme le suivant:

Le premier chapitre est consacré d'abords à un exposé général sur la qualité de service dans le secteur publique. Ensuite, dans le deuxième chapitre nous avons mis l'accent sur les processus stochastiques qui servent comme un outil mathématique de base pour résoudre les problèmes d'attente. Dans le troisième chapitre nous avons présenté les notions de base exprimant les phénomènes d'attente, et nous avons défini aussi les éléments importants à considérer dans la mesure de performance d'un système de file d'attente. Enfin, dans le dernier chapitre nous avons utilisé la théorie des files d'attente pour la commande optimale d'un carrefour à feux.

Chapitre I :

La qualité de service dans le secteur publique

I.1 Introduction

Aujourd'hui, la qualité constitue l'un des moteurs essentiels de la compétitivité moderne, elle est un enjeu stratégique majeur dans une économie de plus en plus globale. La nécessité pour l'homme de résoudre les problèmes de plus en plus complexes, l'ouverture des marchés aux produits et aux services, la croissance menaçante des moyens de communication et d'échange a contraint l'homme à intégrer une démarche qualité au développement des produits ou des services qu'il conçoit.

Pour bien comprendre la philosophie de la qualité ce premier chapitre a pour but d'éclaircir les notions de base de la qualité ainsi que la démarche qualité dans le secteur publique.

I.2 Définition de la qualité

La définition la plus couramment retenue de la qualité est celle que lui confère l'International Organisation for Standardisation (ISO) dans la norme ISO 8402 : « la qualité est l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confèrent l'aptitude à satisfaire des besoins exprimés ou implicites ». Qui a été modifie par **:** « Aptitude d'un ensemble de caractéristiques intrinsèques à satisfaire des exigences » dans la norme **ISO 9000 (2000)**. La qualité n'est donc pas la performance maximale, ni le luxe, ni le « haut de gamme », mais juste la réponse « ajustée » et économique à un besoin donné [1].

I.3 Les enjeux de la qualité de service

I.3.1 Enjeu économique :

- Diminution des coûts
- Augmentation de la valeur ajoutée
- Diminution du non qualité

I.3.2 Enjeu par rapport au client :

- Donner confiance et satisfaction aux clients
- Fidéliser les clients actuels et en gagner de nouveau
- Diminuer les réclamations

I.3.3 Enjeu stratégique :

• Améliorer l'image de marque en diminuant les mauvaises références et en augmentant la notoriété.

• Renforcer son avantage par rapport à la concurrence

• Se développer

I.3.4 Enjeu humain :

• Mise en valeur du travail du personnel et augmentation de la motivation vis à vis d'un travail bien fait.

• Améliorer les relations internes, moins de désordres et plus de prévention.

• Mobilisation plus importante par un travail bien fait.

1.4 Les dimensions de la qualité

La qualité peut être un facteur stimulant pour l'entreprise, car elle peut être source de beaucoup de changement pour son mode de fonctionnement. C'est à dire l'instauration du management de la qualité, contraint l'entreprise à changer et à innover sa politique. L'entreprise peut :

- Améliorer les rapports humains entre les différents groupes des personnes (dimension humaine).

- Innover les capacités de production (dimension technique).

- Réduire les coûts de la non-qualité (dimension économique)…etc.

1.4.1- La dimension humaine

La dimension humaine de la qualité est importante, pour obtenir cette qualité qui est à la fois des principes, des outils, des méthodes, mais aussi et surtout un nouvel état d'esprit qui fait stimuler l'intelligence et la créativité de tous personnels. Ceci s'inscrit d'ailleurs dans la perspective du développement de la participation.

1.4.2- La dimension technique

La qualité doit être d'abord le soucie de proposer un produit disposant des caractéristiques propres à satisfaire les attentes, exprimées ou implicites, des clients pour mieux les attacher et conquérir de nouveaux marchés.

1.4.3- La dimension économique

Il s'agit en effet de satisfaire les besoins au moindre coût, et donc de s'organiser pour réduire au minimum les dysfonctionnements. Il n'est jamais inutile de rappeler le gisement potentielle de profit constitué par l'usine « fantôme » : le coût invisible de l'ensemble des défaillances entraînées par un défaut de qualité (Gaspillage, retouches, rebuts…) coût d'obtention de la qualité.

1.4.4- La dimension organisationnelle

Par l'introduction du système qualité (qualité totale) l'organisation de différentes fonctions deviennent plus productif et mieux gérer. Donc la fonction qualité doit trouver une place dans l'organigramme de l'entreprise et les responsabilités doivent être clairement définies

I.5 Le cycle de la qualité de service

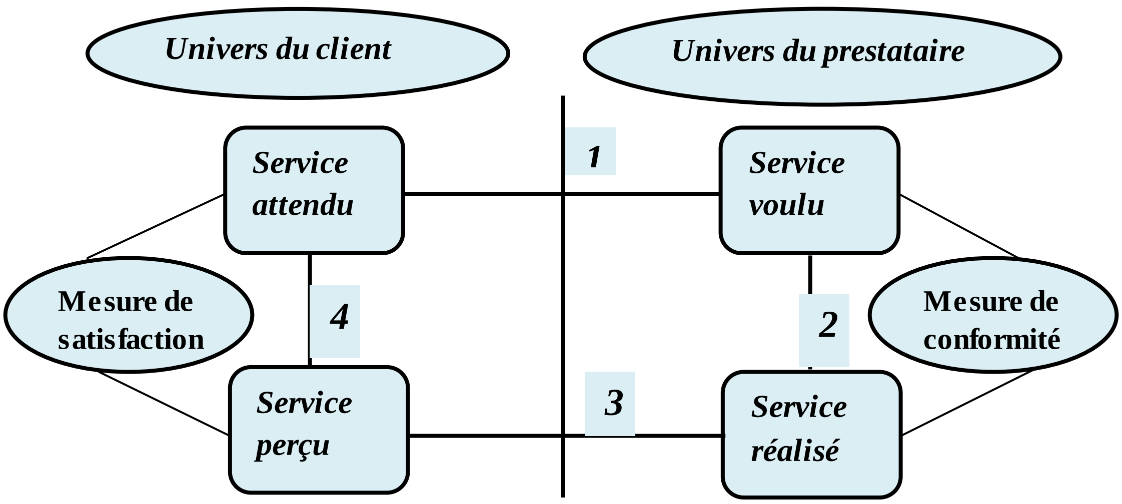


Figure I.1 : Le cycle de la qualité de service

Le modèle définit quatre types de services

• **Le service attendu par le client :** est basé sur ses besoins plus ou moins exprimés et sur son expérience. " C'est l'expression d'un souhait dans un système de référence ".

C'est celui que le client trouve " raisonnable " d'espérer à cet instant, ici et à ce prix. Il s'agit du niveau de qualité explicitement ou implicitement recherché par le client.

• **Le service perçu par le client** : qui est une perception subjective de la situation et non l'évaluation objective du service. La perception de la qualité réalisée dépend pour le client de son expérience personnelle du service ou des prestations qui lui sont associés.

• **Le service voulu :** est l'offre que défini l'entreprise à partir des besoins du client. C'est ce qu'elle souhaite réellement offrir au client. Il prend en compte les contraintes économiques, les savoir-faire, la concurrence... C'est un choix de la direction générale. L'entreprise définit les objectifs de service à rendre en considérant les facteurs suivants :

♦ Une définition concise du service de référence (fréquence, temps d'attente maximum, ponctualité, traitement des réclamations clients, …)

♦ Un niveau de réalisation c'est à dire un pourcentage (estimé ou calculé) de client bénéficiant du service de référence).

♦ Un seuil de non acceptabilité (quand ce seuil est atteint le service est considéré comme non rendu).

• **Le service réalisé** effectivement par l'entreprise. C'est-à-dire " le relevé précis, objectif, de la réalisation du service sur le terrain et au quotidien. La qualité de service évolue dans 2 univers :

♦ Celui du client

♦ Celui de l'entreprise

I.6 Les normes ISO

Les grands domaines de certification internationale sont le Management de la Qualité (Normes **ISO 9000**) et le Management environnemental (Normes **ISO14000)**.

La norme ISO peut être appliquée à tout organisme grand ou petit, quel que soit son produit, ou service, que l'organisme soit une entreprise commerciale, une administration publique ou un département.

I.6.1 Définition de la certification

La certification est une démarche volontaire par laquelle une tierce partie donne une assurance écrite qu'un produit, une organisation, un processus, un service ou les compétences d'une personne sont conformes à des exigences spécifiées. Elle est distincte des autres systèmes de preuves de conformité telles les déclarations du fournisseur, les rapports d'essais d'un laboratoire ou le rapport d'un organisme de contrôle. La certification s'appuie sur des résultats d'essais, de contrôles et d'audits, et donne confiance au client en raison de l'intervention systématique d'un organisme tiers indépendant.

L'ISO donne la définition suivante : « Procédure par laquelle une tierce partie donne une assurance écrite qu'un produit, un processus ou un service est conforme aux exigences spécifiées dans un référentiel ».

La certification est donc le moyen d'attester, par l'intermédiaire d'un certificateur, de l'aptitude d'un organisme à fournir un service, un produit ou un système conforme aux exigences définies dans un cahier de charges.

I.6.2 Types de certification

Il existe différents types de certification :

- La certification de produits et de services atteste que des produits répondent à des caractéristiques de sécurité, d'aptitude à l'usage, d'interchangeabilité définies dans un référentiel. La marque NF atteste la conformité à des caractéristiques définies dans des normes et des spécifications complémentaires aux normes lorsqu'elles sont demandées par le marché.

- La certification d'entreprise démontre la conformité du système qualité de l'entreprise à l'un des trois modèles normalisés par l'ISO (ISO 9001, 9002 ou 9003). Les différents référentiels ne sont pas attachés à un niveau de performance du produit. Ils prennent différemment en compte les processus allant de la conception d'un produit ou d'un service, au soutien après la vente.

I.6.3 Historique et présentation des normes ISO

Les normes ISO sont des normes internationales, définies par l'Organisme International de Standardisation. Elles s'adressent à tous types d'entreprises (agroalimentaires ou non) fabriquant des produits ou proposant des services.

Crées en 1987 à la demande des industriels, les normes ISO 9000 concernent les systèmes de management de la qualité. Ceux-ci souhaitaient un référentiel leur permettant de faire reconnaître leur système de gestion de la qualité auprès de leurs différents clients, notamment du point de vue du respect des cahiers de charges. Elles ont été révisées en 1994 (ISO 9001, 9002 et 9003) puis en 2000.

La principale évolution entre la version 1994 et la version 2000 est le cadre du référentiel. Ainsi, il ne s'applique plus au produit ou service en lui-même mais couvre l'ensemble du système de management de la qualité.

Ce référentiel n'est pas imposé par la législation : c'est une démarche volontaire de l'entreprise (ou demandée par ses clients) de faire reconnaître la conformité de ses méthodes de travail à un référentiel ISO.

La forme abrégée du nom de l'organisation est toujours « ISO », pour toutes les langues, dont la traduction en français correspond à « organisation internationale de standardisation (ou de normalisation).

La Norme ISO 9000 regroupe un ensemble de normes qui appartiennent à la même famille :

	Référence	Titre
Comprendre	ISO 9000	Systèmes de management de la qualité – principes essentiels et vocabulaire
Construire	ISO 9004	Systèmes de management de la qualité – lignes directrices pour l'amélioration des performances
Démontrer	ISO 9001	Systèmes de management de la qualité – exigences
Améliorer	ISO 19011	Audit des systèmes de management de la qualité et de l'environnement

Tableau I.1 : Les normes qui appartiennent à la même famille ISO 9000

I.6.4 Interaction entre les normes

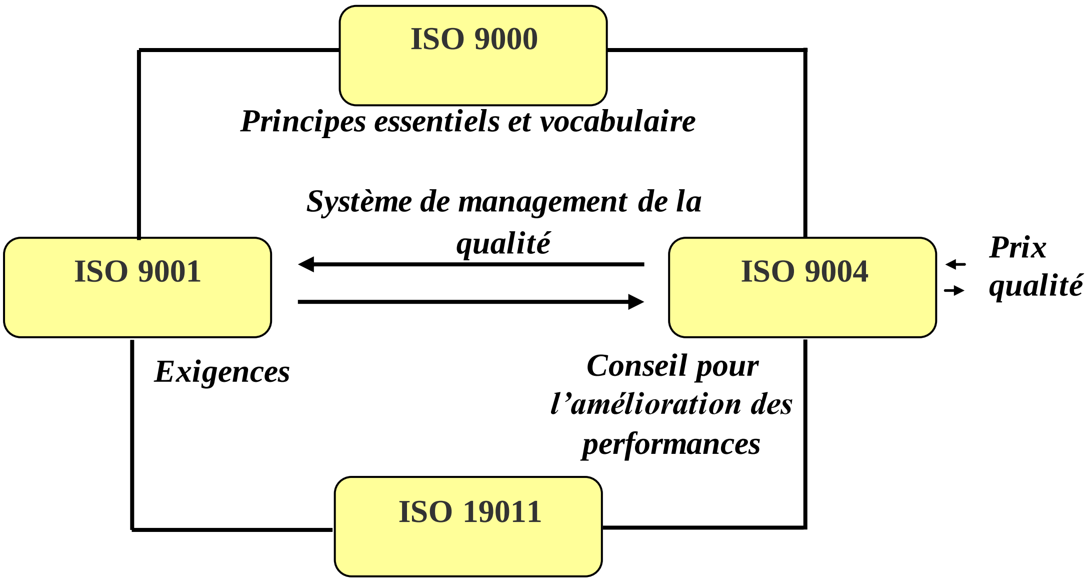


Figure I.2 : Interaction entre les normes de la qualité

Les organismes peuvent également choisir d'être certifiés par d'autres référentiels que l'ISO, qui est une certification du système, ils peuvent choisir une certification de service qui se fait sur la base d'un référentiel sur mesure dédié à un type de structures spécifique, et qui vise à garantir le respect d'engagement de qualité sur un ensemble de prestations. Un référentiel de service comporte des exigences sur l'offre de service et sur les moyens à mettre en œuvre pour en garantir la fiabilité. Nous pouvons citer comme référentiel pour la certification de service : QUALICERT, QUALIPREF, et pour qu'ils puissent servir de support à des certifications, ils doivent être validé par un organisme accrédité.

Mais l'ISO garde l'avantage d'amener un travail plus approfondi sur l'organisation ainsi qu'une crédibilité internationale, mais nous ne pouvons pas faire une hiérarchie entre les deux, car chacune dépend du type de mission et de public, par exemple dans le privé, l'ISO est plus répandu, dans l'administration, la certification de service est sans doute le meilleur moyen de susciter la confiance des usagers, mais l'ISO reste quand même privilégié même dans le public.

I.7 La démarche qualité dans l'entreprise de service

La qualité, c'est la conformité du produit ou service rendu à la promesse marketing et/ou à l'engagement de service.

" L'important ce n'est pas tant la recherche de la perfection ou de travailler plus, que de dire à nos clients ce que l'on va faire pour eux et s'y tenir ". (Serge Alécian)

La qualité est une exigence contemporaine, et non une option facultative. Elle joue et jouera le rôle de " sélection naturelle " auprès des entreprises et des organisations.
Mais la diversité des organisations, de leur finalité, de leur structure, de leur contexte, nécessite une démarche qualité adaptée à chacune.
Une approche, très modulable, souple, progressive, permet la mise en œuvre d'une démarche qualité cohérente avec la stratégie des organisations ou des entreprises concernées.
Cette approche repose cependant sur quelques " croyances " fondamentales:
1- La qualité de service, ce n'est pas ce qui est luxueux, c'est ce qui est conforme à des engagements de service.

2- La qualité de service perçue par le client, n'est que le reflet de la qualité du management et de l'organisation interne.

3- 80% des causes de non-qualité sont imputables à l'action de l'encadrement, 20% sont à celle des opérationnels.

4- Le coût de la qualité est toujours inférieur au coût de la non-qualité.

5-Agir pour la qualité, ce n'est pas seulement agir sur les processus ou les produits, c'est aussi agir sur le management et les hommes.

6- La qualité est l'affaire de tous, dans un rapport gagnant/gagnant.

Ces 6 idées forcent l'intégration de la démarche qualité dans un processus de management global, et constituent la toile de fond d'une approche qualité.

Les trois âges de la qualité, au sein d'une organisation professionnelle

1- La prise de conscience et la "mobilisation".

2- La maîtrise des engagements ou notion de Garantie de Services.

3- L'amélioration du service ou optimisation.

I.8 Les documents de base d'un système qualité

I.8.1 Le manuel qualité

Ce manuel doit décrire les dispositions générales contribuant à la qualité, applicable à toutes les activités de l'organisme. Il fait référence, s'il y a lieu, aux procédures ou autres documents applicables au niveau de l'organisme. Pour les organismes à produits multiples ou à établissement multiples, un manuel qualité général peut être complété par des manuels qualité établis au niveau des différents secteurs; ces manuels font alors référence aux seules procédures et autres documents applicables à ces secteurs. Le manuel qualité est établi par l'organisme essentiellement pour son usage interne.

I.8.2 Le plan qualité

Le plan qualité est établi par l'organisme pour décrire les dispositions spécifiques qu'il prend pour l'obtention de la qualité pour un produit particulier. Lorsque l'organisme dispose d'une manuelle qualité, le plan qualité peut se limiter à la description ou aux références des dispositions spécifiques, pour ce produit particulier, pour les différents secteurs concernés. Le plan qualité est établi par l'organisme essentiellement pour son usage interne.

I.8.3 Le manuel assurance qualité

Le manuel assurance qualité est établi par l'entreprise pour décrire les dispositions générales qu'elle prend en application d'un référentiel contractuel, en particulier les normes NF EN29- 001, 29-002, 29-003 pour l'assurance de la qualité. Il peut être constitué d'extrait du manuel qualité. C'est un élément de base dans les relations clients-fournisseurs ou dans le cadre de la certification du système qualité de l'entreprise par tierce partie.

I.8.4 Le plan assurance qualité

Le Plan assurance qualité est établi par l'organisme pour décrire les dispositions spécifiques qu'il prend en matière de l'assurance de la qualité pour un produit particulier et répondre aux exigences contractuelles de l'assurance de la qualité exprimée par son client. Ces dispositions spécifiques prévalent sur toutes autres dispositions. Un plan d'assurance de la qualité peut intégrer des plans d'assurance qualité fournis par des cotraitants ou des sous-traitants.

I.9 Le Mangement par la Qualité Totale (TQM)

La direction et le fonctionnement d'un organisme avec succès nécessitent de l'orienter et de le contrôler méthodiquement et en transparence. Le succès peut résulter de la mise en œuvre et l'entretien d'un système de management conçu pour une amélioration continue des performances tout en répondant aux besoins de toutes les parties intéressées. Le management par la qualité est une composante du management d'un organisme. Il se définit comme l'ensemble des activités permettant d'orienter et de contrôler un organisme en matière de qualité.

Ceci inclut généralement l'établissement d'une politique qualité et d'objective qualité, la maîtrise de la qualité, l'assurance de la qualité et l'amélioration de la qualité.

Selon les principes du management par la qualité, les entreprises doivent définir leurs missions, leurs visions et leurs objectifs. Elles doivent viser une amélioration constante du produit ou du service qu'elles offrent à leurs clients. Elles ne peuvent y parvenir sans maintenir un niveau élevé de motivation et de satisfaction des salariés de l'entreprise. C'est après avoir défini leur mission que les entreprises peuvent se mettre à l'œuvre pour la réaliser.

Le secteur des services se distingue des autres secteurs économiques en ce sens que le personnel en est le principal intrant. Il est essentiel que celui-ci développe et renouvelle ses connaissances afin d'optimiser le potentiel d'emplois qu'offrent les services. Quatre-vingt pour cent des entreprises européennes en forte croissance considèrent que la formation du personnel a joué un rôle décisif dans leur réussite. La formation deviendra de plus en plus importante à l'avenir. Si les personnes constituent la principale ressource, la recherche de qualité doit être centrée sur elles. On ne peut pas demander aux salariés de fournir des services de qualité si leur environnement de travail est inadéquat.

Les dirigeants doivent constamment apprendre de leur personnel, autant que des clients et des concurrents. La démarche qualité totale doit se déployer à tous les niveaux de l'entreprise, en commençant par un engagement total au sommet. Tous les obstacles organisationnels et physiques au travail en équipe doivent être éliminés, en particulier les évaluations de la performance.

L'essence des pratiques de management par la qualité totale réside dans «le management des systèmes et processus». Cette méthode préconise que les entreprises doivent étudier et comprendre avec toujours plus de précision leurs processus de production ou de prestation de services. Si l'entreprise veut diviser son système en différents blocs à des fins d'analyse, elle se doit d'étudier aussi les «clients internes» qui mettent en œuvre les processus demandés.

Le management par la qualité totale renferme trois notions fondamentales. La première est que les personnes sont l'atout le plus précieux de toute organisation ou entreprise. Les organisations qui aident et soutiennent leur personnel à épanouir leur potentiel peuvent engranger des bénéfices considérables. Ce devrait être la mission principale du management.

Deuxièmement, les organisations cherchent à s'améliorer et à valoriser leurs processus de travail sur la base de données objective. L'analyse statistique est donc un outil essentiel dans la démarche TQM. Troisièmement, le TQM vise avant tout la satisfaction du client, qui est l'utilisateur final du service ou du produit.

Le management par la qualité exige de ce fait la maîtrise des huit principes basique de la démarche qualité pour mener l'organisme vers les meilleures performances, à savoir :

1. Orientation client

La compréhension et la satisfaction des attentes et besoins des clients représentent aujourd'hui une orientation fondamentale de la démarche qualité. Elles conditionnent le développement et la pérennité des entreprises. L'écoute et l'attention aux clients doivent être présentes à toutes les étapes de la vie des produits et services, de leur conception à leur utilisation par les clients, et à tous les niveaux d'une entreprise. Cette préoccupation décisive oriente non seulement l'activité quotidienne des professionnels mais également les choix stratégiques essentiels, notamment le développement de nouveaux services.

2. leadership

Dans tout système de management de la qualité, la direction doit déterminer clairement ses orientations stratégiques et créer les conditions pour que le personnel puisse pleinement s'impliquer. Pour cela elle doit monter l'exemple et son réel engagement, définir des objectifs motivants et créer des valeurs partagées.

3. Implication du personnel

Le personnel est le cœur même d'une entreprise et donc l'un des maillons principal pour tout système de management de la qualité. Son implication est indispensable pour qu'une entreprise puisse progresser. Il est important de faire comprendre à chacun son rôle et son importance, de les responsabiliser, de mettre en place des méthodes pour les valoriser dans leur implication.

4. Le management de la qualité par approche système

Comprendre et gérer l'entreprise comme un système de processus indépendants en vue d'un objectif donné permet d'améliorer son efficacité et son efficience. Ce principe permet de clarifier le fonctionnement de l'entreprise, de mettre à jour et de supprimer les activités qui sont souvent source de dysfonctionnements.

5. Approche factuelle pour la prise de décision

Les décisions efficaces se fondent sur l'analyse de données et d'informations. Pour pouvoir prendre les bonnes décisions, il faut pouvoir s'appuyer sur des informations fiables. Ces informations doivent donc être disponibles et sous une forme permettant leur analyse et leur compréhension.

6. Relations mutuellement bénéfiques avec les fournisseurs

Une entreprise et ses fournisseurs sont interdépendants et des relations mutuellement bénéfiques permettront d'augmenter leurs capacités à créer de la valeur. Pour cela, il est nécessaire de comprendre les intérêts des partenaires, de définir clairement leurs obligations et d'évaluer régulièrement leurs performances et de les intégrer dans les processus.

7. L'amélioration continue

L'amélioration continue d'un système de management de la qualité consiste à augmenter la performance interne et la satisfaction des clients .Cela comprend ente autre :

- analyse des résultats pour identifier les pistes d'amélioration,
- établissement des objectifs,
- recherche et mise en œuvre des actions d'amélioration,
- formalisation des changements.

Le principe de l'amélioration continue est souvent représenté par un cycle d'action, appeler roue de Deming (figure I.3) afin de désigner les quatre temps suivants :

- **Planifier** : Définir les objectifs à atteindre.
- **Faire** : il s'agit de la mise en œuvre des actions correctives.
- **Mesurer/vérifier** : Vérifier l'atteinte des objectifs fixés.
- **Agir** : Prendre des mesures préventives.

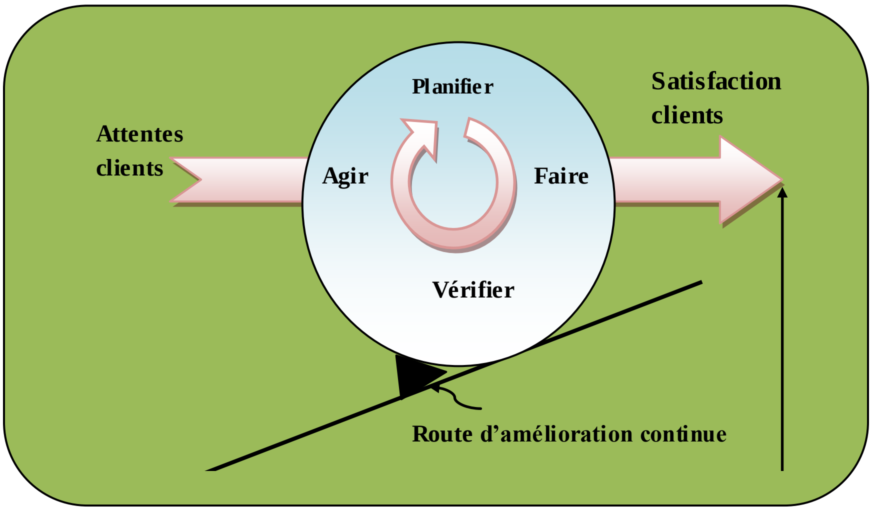


Figure I.3 : Roue de Deming

8. Approche processus

Pour qu'un organisme fonctionne de manière efficace, il doit identifier et gérer de nombreuses activités corrélées. Toute activité utilisant des ressources et gérée de manière à permettre la transformation d'éléments d'entrée en éléments de sortie, peut être considérée comme un processus. L'élément de sortie d'un processus constitue souvent l'élément d'entrée du processus suivant.

« **L'approche processus** » désigne l'application d'un système de processus au sein d'un organisme, ainsi que l'identification, les interactions et le management de ces processus.

Lorsqu'elle est utilisée dans un système de management par la qualité, cette approche souligne l'importance :

a) de comprendre et de satisfaire les exigences ;

b) de considérer les processus en termes de valeur ajoutée ;

c) de mesurer la performance et l'efficacité des processus ;

d) d'améliorer en permanence des processus sur la base de mesures objectives.

Le modèle de système de management par la qualité basé sur les processus présenté en Figure I.4 illustre les relations entre les processus. Cette figure montre le rôle significatif joué par les clients lors de la définition des exigences en tant qu'éléments d'entrée. La surveillance de la satisfaction des clients exige l'évaluation des informations concernant la perception des clients sur le niveau de réponse de l'organisme de leurs exigences. Ce modèle couvre toutes les exigences de la présente norme internationale mais ne présente pas les processus à un niveau détaillé.

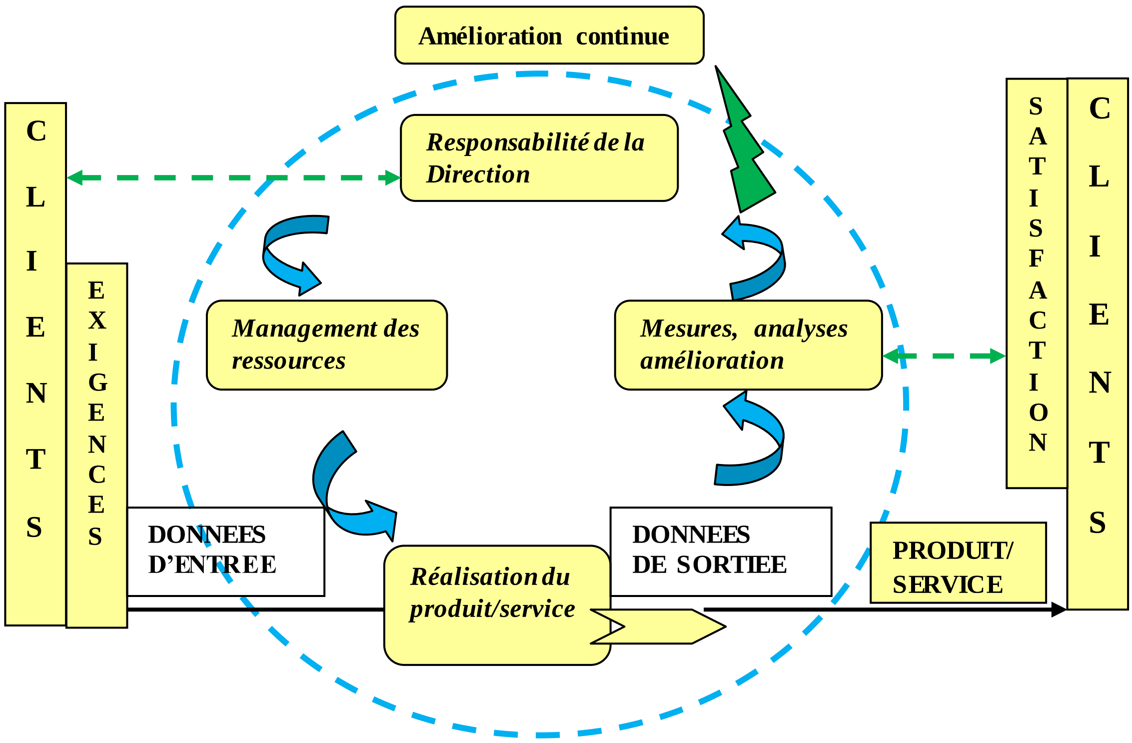


Figure I.4: Modèle d'un système de management par la qualité basé sur les processus [Michel Perigod 1990]

I.10 Conclusion

Dans ce chapitre, nous avons présenté d'abord la notion da la qualité dans le secteur publique, ensuite nous avons défini les principes de management par la qualité totale.

Aujourd'hui plus qu'hier et partout dans le monde, la concurrence se fait plus vive en effet, le progrès de la productivité et le rythme de l'innovation génère une offre chaque jour plus importante et variée et le consommateur de bien, ou l'utilisateur de services peut désormais se montrer plus exigeant. De ce fait, les entreprises sont donc amenées à adopter des démarches qualités en intégrant, dans leurs stratégies les techniques modernes de management de la qualité. La stratégie qualité constitue alors pour les dirigeants de l'entreprise un impératif pour le maintien de la compétitivité sur les marchés aussi bien qu'interne qu'externe.

Chapitre II :

Généralités sur les processus stochastiques

II.1 Introduction

Nous abordons dans ce chapitre les processus à la base de l'étude de tels systèmes d'attente qui sont des processus stochastiques. Ces processus permettent de modéliser des systèmes dont le comportement n'est que partiellement prévisible. La théorie est fondée sur le calcul des probabilités et les statistiques. Les domaines d'application sont très divers surtout dans la modélisation et la gestion du trafic dans les réseaux de transport (carrefours à feux) et plus généralement la gestion des systèmes techniques complexes soumis à des perturbations aléatoires.

Nous exposons dans un premier temps les notions de base concernant les processus stochastiques, puis dans un second temps nous metterons l'accent sur les notions théoriques fondamentales concernant le processus markovienne, processus de Poisson et processus de naissance et de mort qui sont des préliminaires nécessaires à l'étude des systèmes de files d'attente.

II.2 Notations et définitions

L'étude mathématique d'un système d'attente, se fait le plus souvent par l'introduction d'un processus stochastique approprie. Généralement, on qualifie de processus stochastique tout phénomène d'évolution temporelle dont l'analyse peut être soumise au calcul des probabilités. Du point de vue de l'observation, un processus stochastique est constitué par l'ensemble de ses réalisations. Une réalisation est obtenue par l'expérience qui consiste à enregistrer une suite d'événements au cours du temps. Le caractère aléatoire de l'évolution se montre par le fait que la répétition de l'expérience conduit à une autre séquence temporelle. Les exemples sont innombrables dans la vie quotidienne comme l'arrivé des voiture dans un carrefour à feux. [9].

Définition : Un processus stochastique peut être défini comme un « ensemble des phénomènes se produisent dans le temps et régi par le hasard. En effet, stochastique signifie fruit de hasard qui comporte donc la présence d'une variable aléatoire ».

Un processus stochastique est ainsi un « ensemble des phénomènes produit par le hasard dans le temps, et dont l'évolution peut être décrite à l'aide de variable aléatoire »[11].

Mathématiquement, un processus stochastique $\{x_t, t \in T\}$ est une collection de variables aléatoires indexées par un paramètre t et définies sur un même espace de probabilités (Ω, F, P).

La variable x_t représente l´état du processus au temps t et l'ensemble de toutes les valeurs possibles pour cette variable est appelée l'espace des états du processus et sera noté N.

II.3 Classification des processus stochastiques

La classification d'un processus stochastique dépend de trois quantités :

- l'espace d'état (N),
- l'indice t,
- la dépendance statistique entre les variables aléatoires pour différentes valeurs de l'indice t.

Lorsque la variable x_t peut prendre un nombre fini ou dénombrable de valeurs, le processus est dit à espace d'état discret. Si au contraire, ses valeurs appartiennent à un espace d'état continue, il est dit à espace d'état continue.

II.4 Chaine de Markov

Après l'explication de la notion d'un processus stochastique, nous nous intéressons maintenant à la présentation de la chaîne de Markov qui sert comme l'outil le plus important dans l'analyse des phénomènes d'attente .Nous abordons dans le premier temps la représentation graphique d'une chaîne de Markov, puis les probabilités et les états possible pour une chaine de Markov.

Une suite $\{X_t, t = 0,1,2,3, \dots\}$ est dite possédant une propriété markovienne ou est une chaîne de Markov, si pour chaque instant t et pour chaque état de X on a :

$$P\ \{X_{t+1} = j / X_0 = a, X_1 = b, \dots, X_t = i\} = P\ \{X_{t+1} = j / X_t = i\} \qquad \text{(II.1)}$$

Cette propriété markovienne signifie qu'étant donné l'ensemble des états passés et présentes du système, la probabilité d'un état futur quelconque de ce système est indépendante de son état actuel ; c'est-à-dire l'état courant résume, à lui seul, tout l'historique du système susceptible d'influencer son évolution future. On parlera dans ce cas de processus sans mémoire.

Les probabilités conditionnelles $P\{X_{t+1} = j / X_t = i\}$ sont appelées probabilités de transition et sont notées tout simplement $P_{i,j}$ constituant ainsi une matrice carrée d'ordre n.

P_{02}

$$P_{i,j} = \begin{vmatrix} P_{00} & P_{01} \ldots\ldots\ldots\ldots\ldots & P_{0n} \\ P_{10} & P_{11} \ldots\ldots\ldots\ldots\ldots & P_{1n} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\ P_{i0} & P_{i1} \ldots\ldots P_{ij} \ldots\ldots\ldots & P_{in} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\ P_{n0} & P_{n1} \ldots\ldots\ldots\ldots\ldots & P_{nn} \end{vmatrix}$$

II.4.1 Chaîne de Markov homogène

Une "chaîne de Markov homogène" est une chaîne telle que la probabilité qu'elle a pour passer dans un certain état au $n^{ème}$ état soit indépendante du temps. En d'autres termes, la loi de probabilité caractérisant la prochaine étape ne dépend pas du temps (de l'étape précédente), et en tout temps la loi de probabilité à tout moment de la chaîne est toujours la même pour caractériser la transition à l'étape en cours.

II.4.2 Matrice stochastique

Une matrice carrée $P = (P_{i,j})$ est stochastique si :

1) ses éléments sont non négatifs : $P_{i,j} \geq 0$ pour tout i et j ;

2) la somme des éléments de chacune de ses lignes est égale à 1 :

$$\sum_{j=1}^{n} P_{i,j} = 1, \text{ pour tout } i. \qquad \text{(II.2)}.$$

II.4.3 La représentation graphique d'une chaîne de Markov

Les chaînes de Markov peuvent être représentées graphiquement sous la forme d'un graphe orienté G ayant pour sommet les points i et pour arêtes les couples orientés (i, j). Nous associons alors à chaque composant un arc orienté et sa probabilité de transition.

Exemple:

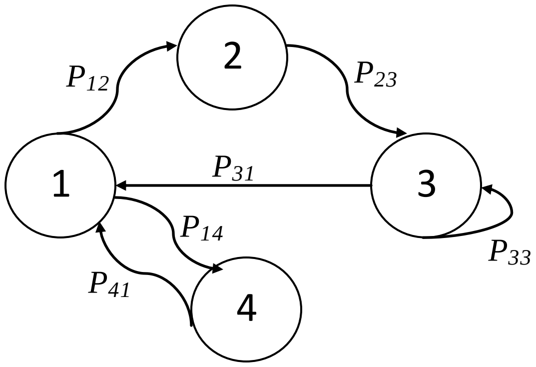


Figure II.1 : exemple d'une représentation graphique d'une chaîne de Markov

Dans cet exemple, les seules transitions permises par les 4 états (matrice 4x4) sont celles indiquées par les flèches. Ce qui fait que la matrice de transition s'écrit alors :

$$P_{i,j} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & P_{12} & 0 & P_{14} \\ 0 & 0 & P_{23} & 0 \\ P_{31} & 0 & P_{33} & 0 \\ P_{41} & 0 & 0 & 0 \end{bmatrix}$$

II.4.4 Les probabilités d'états pour une chaîne de Markov

Après la représentation graphique d'une chaîne de Markov nous nous intéressons maintenant au calcul des probabilités d'états. Désignons par $P(n)$ le vecteur stochastique défini par $P(X_n = j), j = 1,2,3 \ldots$ Ainsi $P(0)$ sera la distribution initial, alors on peut déduire que :

$$P(n) = P(n-1)[P_{i,j}], \quad (II.3).$$

De même, on peut écrire :

$$P(n-1) = P(n-2)[P_{i,j}] \quad (II.4),$$

Et ainsi de suite jusqu'à :

$$P(2) = P(1)[P_{i,j}] \text{ ,et}$$

$$P(1) = P(0)[P_{i,j}] .$$

En substituant $P(1)$, par sa valeur, dans $P(2)$, $P(2)$ dans $P(3)$,…etc, on pourra donc écrire :

$$P(n) = P(0)[P_{i,j}]^n \quad (II.5).$$

Ce qui signifie que l'état du système après la $n^{ème}$ transition ne dépend que de la distribution initial et de la matrice de transition.

II.4.5 Définition des états possible dans une chaîne de Markov

1) Etat accessible : On appelle état accessible à partir de l'état E_i tout état E_j d'une chaine de Markov tel que $P_{i,j}(q)^n > 0$ au moins pour certaines valeurs de q. Deux états mutuellement accessibles l'un à partir de l'autre sont dits communicants.

2) Etat récurent : Désignons par τ_{mn} le temps de premier passage de l'état m à l'état n dans une chaîne de Markov discrète :

$\tau_{mn} = \min_{k>1}\{k : X_k = n / X_0 = m\} = k_{k\geq 1}\{k : X_{k+r} = n / X_r = m\}$

(propriété du à l'homogénéité de la chaine).

L'état m est dit état récurent si $P[\tau_{mn} < \infty] = 1$

Les états récurent sont dits aussi persistants.

3) m est dit transitoire si $P[\tau_{mn} < \infty] < 1$

***4)* Etats périodiques** : Les états périodiques sont des états qui ne sont atteints que périodiquement.

Remarque : Si tous les états d'une chaîne sont communicants mutuellement, alors elle est dite irréductible.

II.4.6 Propriétés d'ergodicité d'une chaîne de Markov

Si $\lim M^m = T$ quand $m \to +\infty$ où T est une matrice stochastique ne possédant aucune élément nul, on dit que le système est ergodique, ou qu'il est stable en probabilité, ou encore qu'il possède un régime permanent, et M est dite matrice ergodique.

II.5 Processus de naissance et de mort

Pour pouvoir étudier ce processus, on devrait d'abord envisager l'étude des processus suivants : processus de naissance, processus de Poisson, et processus de mort. Nous étudions d'abord le processus de naissance puis nous intéressons au processus de Poisson qui est fréquemment utilisée pour modéliser les occurrences d'un événement pouvant survenir à tout instant avec une probabilité constante et indépendamment des occurrences passées, ensuite nous étudierons le processus de mort, et nous nous intéresserons enfin au cas générale dans lequel en considère l'évolution d'une population qui connaît à la fois des naissances et des morts.

II.5.1 Processus de naissance

On qualifie un processus de processus de naissance, si ce processus est caractérisé par l'apparition d'un individu, au sein d'une population, selon une certaine loi. L'exemple le plus courant dans la vie quotidienne est l'arrivé des voitures dans un carrefour à feux ou une personne devant un guichet.

Un processus de naissance est dit homogène, si la probabilité d'apparition d'un individu pendant l'intervalle Δt, sachant qu'il existe déjà k individus au sein du système, est donnée par :

$$\lambda_k \Delta t + 0(\Delta t) \quad \text{où} \quad 0(\Delta t) \text{ est un infiniment petit.}$$

Si on écarte le cas où deux apparitions simultanées peuvent avoir lieu en même temps, alors on pourrait dire que la probabilité qu'il n'y ait aucune apparition dans l'intervalle Δt sachant qu'il y en a eu déjà k est égale à :

$$1 - \lambda_k \Delta t + 0(\Delta t).$$

Notons par N_t le nombre d'apparitions durant l'intervalle $[0, t]$ et soit $P_n(t)$ la probabilité qu'à l'instant t l'effectif soit n.

D'où $P_{i,j}(t)$ est la probabilité qu'à l'instant t l'effectif est j sachant qu'il y avait déjà i individus dans le système.

$P_{i,j}(t) = 0$ si $j < i$ (On suppose qu'il n'y a eu aucune disparition).

D'autre part $P_n(t) = P_{0,n}(t) = P_{1,n+1}(t) = \cdots$

Le processus ainsi décrit pourrait-être schématisé par le graphe suivant :

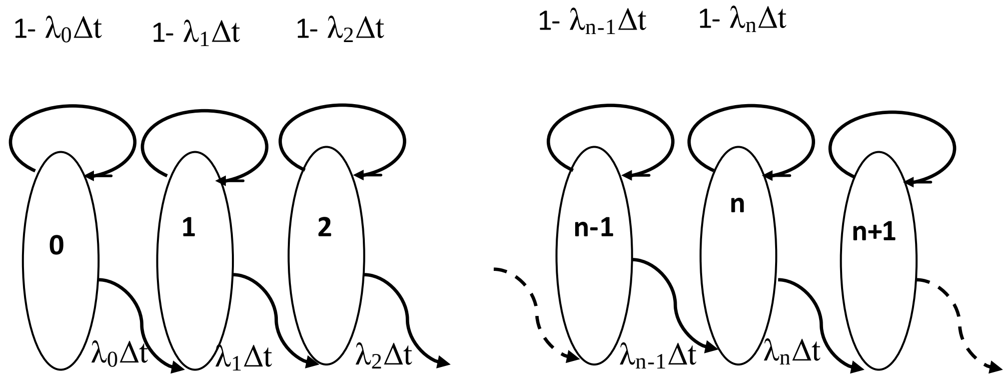


Figure II.2: processus de naissance

La figure (II.2) montre que la possibilité d'être à l'état n à la date i, il y a deux possibilités :

• Le système était dans l'état $n-1$ à l'instant $\mathrm{i}-1$ avec une probabilité :

$\mathrm{P}_{\mathrm{n}-1}(\mathrm{i}-1)$ et il y a eu apparition d'un individu en $\Delta\mathrm{t}$ avec une probabilité : $\lambda_{\mathrm{n}-1}\Delta\mathrm{t}$,

• Ou alors le système était déjà dans l'état n à l'instant $\mathrm{i}-1$ avec une probabilité

$\mathrm{P}_{\mathrm{n}-1}(\mathrm{i}-1)$ et il n'y a pas eu apparition d'un individu en $\Delta\mathrm{t}$ avec une probabilité :

$$1-\lambda_{\mathrm{n}-1}$$

I.5.1.1 La matrice de transition pour un processus de naissance

$M = [\, P_{i,j}(\Delta t)]$

$$
M = \begin{array}{c|ccccccc|}
 & E_0 & E_1 & E_2 & & E_{n-1} & E_n & E_{n+1} \\
\hline
E_0 & 1-\lambda_0\Delta t & \lambda_0\Delta t & 0 & \dots & 0 & 0 & \dots \\
E_1 & 0 & 1-\lambda_1\Delta t & \lambda_1\Delta t & \dots & 0 & 0 & \dots \\
E_2 & 0 & 0 & 1-\lambda_2\Delta t & \dots & 0 & 0 & \dots \\
 & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
E_{n-1} & 0 & 0 & 0 & \dots & 1-\lambda_{n-1}\Delta t & \lambda_{n-1}\Delta t & \dots \\
E_n & 0 & 0 & 0 & \dots & \dots & 1-\lambda_{n\text{-}}\Delta t & \dots \\
E_{n+1} & \dots & \dots & \dots & \dots & \dots & \dots & \dots
\end{array}
$$

II.5.1.2 Les équations futures du système

Considérons le vecteur des probabilités d'état p_n aux instants t et $t + \Delta t$; d'après la définition da la matrice de transition vue dans le paragraphe des chaines de Markov, on peut écrire :

$$[\, P_0(t+\Delta t),\ P_1(t+\Delta t), P_2(t+\Delta t), \dots,\ P_n(t+\Delta t)]$$

$$= [\, P_0(t),\quad P_1(t),\quad P_2(t), \dots ,\quad P_n(t)].M$$

Et de façon explicite les probabilités d'état du système s'écrivent comme suit :

$$P_0(t+\Delta t) = P_0(t)\,(1-\lambda_0\,\Delta t)$$

$$P_1(t+\Delta t) = P_1(t)\,(1-\lambda_1\,\Delta t) + P_0(t)\ \lambda_0\,\Delta t$$

$$P_2(t+\Delta t) = P_2(t)\,(1-\lambda_2\,\Delta t) + P_1(t)\ \lambda_1\,\Delta t$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.....$$

$$P_n(t+\Delta t) = P_n(t)\,(1-\lambda_n\,\Delta t) + P_{n-1}(t)\ \lambda_{n-1}\,\Delta t \qquad (II.7).$$

Considérons cette dernière relation ; on remarque qu'elle pourrait-être écrite sous la forme :

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = \lambda_{n-1} \; P_{n-1}(t) - \lambda_n \; P_n(t)$$

Si Δt tend vers 0 alors $\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t}$ tend vers $P_n'(t)$; ce qui donne :

$$P_n'(t) = \lambda_{n-1} \; P_{n-1}(t) - \lambda_n \; P_n(t) \qquad \forall n = 1,2, \dots \dots$$

Pour n = 0, on peut déduire que :

$$\frac{\lambda_0 \; P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda_0 \; P_0(t)$$

$$\text{Alors} \quad P_n'(0) = \lambda_0 \; P_0(t) \qquad (\text{II}.8).$$

Si on suppose qu'au début le système était vide, alors $P_0(0) = 1$ et $P_n(0) = 0$

$\forall n \geq 1$

II.5.2 Processus de Poisson

Dans le formalisme des files d'attente, les arrivées des clients sont caractérisées par l'ensemble des instants ou dates d'arrivée de chaque client. Cette collection des dates d'arrivée s'appelle le processus des arrivées. Lorsque les dates d'arrivées sont imprévisibles, elles sont modélisées par des variables aléatoires (processus de Poisson).

Nous nous intéressons ici à la présentation du processus de Poisson, puis en metterons l'accent sur la loi exponentielle qui est le plus utilisé dans la modélisation du processus de service.

On remarque que le processus de Poisson au niveau des carrefours à feux est caractérisé par le mode selon lequel s'effectuent les arrivées au sein de ces systèmes. Ce processus est un cas particulier d'un processus de naissance où $\lambda_k = \lambda$.

Désignons par N_t le nombre d'événements qui peuvent se produire pendant l'intervalle $[0, t]$ et qui vérifient les conditions suivantes :

Condition 01 : le processus ainsi décrit est homogène dans le temps ; donc les probabilités de transitions sont constantes.

Condition 02 : le processus est additif à accroissement indépendants : étant donnés deux intervalles de temps disjoints, le nombre des événements qui se produisent dans l'un est indépendant du nombre des événements qui se produisent dans l'autre.

Condition 03 : la probabilité pour que deux événements se réalisent en même temps est négligeable devant la probabilité de réalisation d'un seul.

Condition 04 : on doit écarter le cas des probabilités de transition nulles, ou égales à 1.

Mathématiquement ces conditions peuvent-être résumées par :

$$P(\, N_{t+\Delta t} - N_t > 0 / N_t = n) = \lambda_t \, \Delta t + 0(\Delta t) \qquad \forall n \in N \ \text{ et } 0 < \lambda < 1.$$

$$P(\, N_{t+\Delta t} - N_t > 1 / N_t = n) = 0(\Delta t)$$

Si Δt tend vers 0, on peut écrire de façon plus simple :

$$P(\, N_{t+\Delta t} - N_t = 1 / N_t = n) = \lambda_t \, \Delta t + 0(\Delta t)$$

$$P(\, N_{t+\Delta t} - N_t = 0 / N_t = n) = 1 - \lambda_t \, \Delta t + 0(\Delta t).$$

Le processus ainsi décrit pourrait-être schématisé par le graphe suivant :

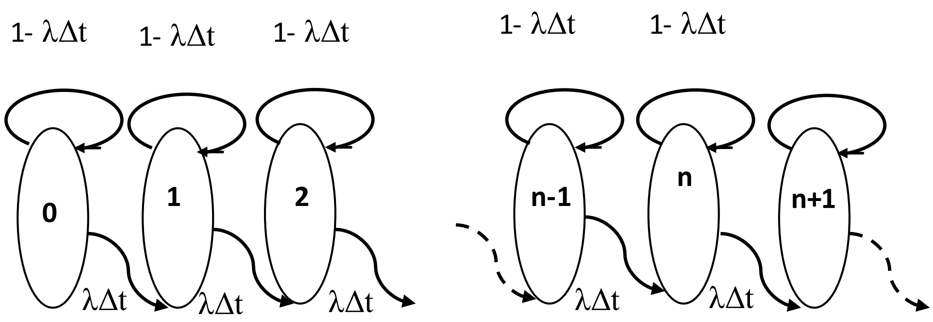


Figure II.3: processus de Poisson

II.5.2.1 La matrice de transition

Le processus de Poisson peut être défini comme un processus markovien dont la matrice de transition est :

$$M = \begin{array}{c|ccccccc|}
 & E_0 & E_1 & E_2 & & E_{n-1} & E_n & E_{n+1} \\
E_0 & 1-\lambda\Delta t & \lambda\Delta t & 0 & \dots & 0 & 0 & \dots \\
E_1 & 0 & 1-\lambda\Delta t & \lambda\Delta t & \dots & 0 & 0 & \dots \\
E_2 & 0 & 0 & 1-\lambda\Delta t & \dots & 0 & 0 & \dots \\
 & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
E_{n-1} & 0 & 0 & 0 & \dots & 1-\lambda\Delta t & \lambda\Delta t & \dots \\
E_n & 0 & 0 & 0 & \dots & \dots & 1-\lambda\Delta t & \dots \\
E_{n+1} & \dots & \dots & \dots & \dots & \dots & \dots & \dots
\end{array}$$

II.5.2.2 Les équations future du système

Pour obtenir les équations du futur, il suffit de multiplier la matrice de transition précédente parle vecteur $[P_0(t), P_1(t), \dots, P_n(t)]$.

On obtient donc :

$$P_0'(t) = -\lambda\, P_0(t) \text{ et } P'n(t) = -\lambda\, P_{n-1}(t) - \lambda\, P_n(t) \qquad \forall n \geq 1.$$

La première équation donne :

$$\frac{P_0'(t)}{P_0(t)} = -\lambda \qquad \Rightarrow \ln P_0'(t) = -\lambda t + c$$

D'où $P_0'(t) = e^{-\lambda t} + c \quad = k\, e^{-\lambda t}$.

Ainsi que $P_0(0) = 1$ donc $k = 1$. Alors :

$$P_0(t) = e^{-\lambda t} \qquad (II.9)$$

Pour $n = 1$ la deuxième équation donne :

$$P_1'(t) = -\lambda\, P_0(t)\; - \lambda\, P_1(t) = \lambda e^{-\lambda t} - \lambda\, P_1(t) \qquad (II.10)$$

On aura donc à résoudre une équation différentielle du type $y' = f(x) - \lambda y$.

En appliquant les règles connues par ce fait, on obtient :

$$P_1(t) = k(t)e^{-\lambda t} \qquad (II.11)$$

Mais $P_1(0) = 0$; donc $k(0) = 0$.

Si on dérive $(II.11)$, on aura :

$$P_1'(t) = k'(t)e^{-\lambda t} - \lambda k(t)e^{-\lambda t}.$$

En identifiant cette expression avec (1.10), on déduit que $k'(t) = \lambda$.

Donc $k\,(t) = \lambda t + c$. Mais :

$$k\,(0) = \lambda * 0 + C = 0 \Rightarrow C = 0 \; et \; k\,(t) = \lambda t.$$

D'où :

$$P_1(t) = \lambda t e^{-\lambda t}$$

Si on agit de la même manière avec les autres équations, on pourrait déduire de façon générale :

$$P_n(t) = \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} \qquad \forall n \geq 1 \;(II.12).$$

Qui donne la probabilité d'occurrence de n événements dans l'intervalle $[0, t]$ et vérifiant les conditions 1 à 4.

II.5.3 La loi exponentielle

Nous nous intéressons maintenant à la loi exponentielle qui est la plus courante dans la modélisation du processus de services dans les files d'attente.

Généralement Une loi exponentielle correspond au modèle suivant :

λ est un réel strictement positif. La fonction f est définie sur $\mathbb{R}$ par :

$$f(t) = \begin{Bmatrix} \lambda e^{-\lambda t} \text{ si } t \geq 0 \\ 0 \text{ si } t < 0 \end{Bmatrix}$$

f est la densité d'une loi de probabilité P appelée loi exponentielle de paramètre λ.

La fonction f est définie, continue, positive sur $[0 ; +\infty[$.

$$\lim_{\to\infty} \int_0^t f(x)dx = 1$$

Si T est une variable aléatoire représentant une durée de vie et suivant une loi exponentielle, alors sa fonction de répartition est donnée par :

$$F(t) = P(T \leq t) = P\,[0\ \ ,t] = \int_0^t f(x)dx = 1 - e^{-\lambda t}.\text{pour } t \geq 0.$$

Le calcul de l'espérance donne $E(T) = \frac{1}{\lambda}$ ce qui correspond à la durée moyenne de vie.

II.5.4 Processus de mort

Ce processus caractérise le phénomène de disparition au sein d'une population. On rencontre ce processus dans les carrefours à feux (départ des voitures).

Remarquant que les hypothèses qui peuvent représenter ce processus sont très voisines de celle déjà vues dans l'étude du processus de naissance. Particulièrement on pose $N_0 = N > 0$ et on suppose que la probabilité pour qu'une disparition ait lieu entre t et $t + \Delta t$ sachant qu'à t il y avait k individus au sein du système, est donnée par $\mu_k \Delta t$ μ_k est le taux de disparition.

Donc on peut écrire :

$$P(\ N_{t+\Delta t} - N_t = -1/\ N_t = k) = \mu_k\, \Delta t + 0(\Delta t) \text{ et}$$

$$P(\ N_{t+\Delta t} - N_t = 0/\ N_t = k) = 1 - \mu_k\, \Delta t + 0(\Delta t)\ .$$

Ainsi que μ_k doit-être positif, sauf pour $k = 0$ où $\mu_0 = 0$.

II.5.5 Processus de naissance et de mort

Ce processus est la fusion des deux processus précédents : le processus de naissance et le processus de mort.

Un processus de naissance et de mort est un processus qui consiste à faire évoluer un système entre une infinité dénombrable ou non dénombrable (processus continu) d'états, le système étant à chaque instant dans un état est un seul. A titre d'exemple : une file d'attente devant un carrefour à feux, les états du système étant le nombre de voitures dans le lieu de service. L'arrivée des voitures peut être considérée comme une naissance et le départ des voitures peut être considéré comme une mort.

Pour analyser le processus de naissance et de mort à espace d'état discret, on fait généralement les hypothèses suivantes :

Hypothèse 1 : le système ne se déplace que vers l'un des états voisins : si le système est dans l'état E_n à l'instant t, il ne pourra passer à l'instant $t + \Delta t$ que dans les états : E_{n+1} et on appelle ce passage « naissance », l'état E_{n-1} désignant le passage « mort », ou l'état de non changement E_n .

Hypothèse 2 : les probabilités de passage d'un état à un autre dépendent de l'état de départ considéré mais non de la date t.

Hypothèse 3 : un événement au plus (naissance ou mort) peut survenir en un instant donné.

On doit supposer que pendant l'intervalle des temps élémentaire Δt, il ne peut se produire qu'un seul évènement : soit une naissance, soit une mort ; car en réalité, il ne s'agit que d'un seul processus.

Désignons par N_t l'effectif au sein du système à la date t. Les probabilités de transitions $p_{i,j}(\Delta t) = p(N_{t+\Delta t} = j / N_t = i)$ doivent vérifier les conditions suivantes :

$$p_{i,i+1}(\Delta t) = \lambda_i \Delta t + 0(\Delta t)$$

$$p_{i,i-1}(\Delta t) = \mu_i \Delta t + 0(\Delta t)$$

$$p_{i,i}(\Delta t) = 1 - (\lambda_i + \mu_i)\,\Delta t + 0(\Delta t)$$

$$p_{i,j}(0) = \delta_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si non} \end{cases}$$

Les λ_i et μ_i e sont strictement positifs, sauf $\mu_0 = 0$ où λ_i est le taux de naissance si i individus existent déjà dans le système, et μ_i est le taux de mort au sein du même système.

Le graphe de transition correspondant à ce système sera :

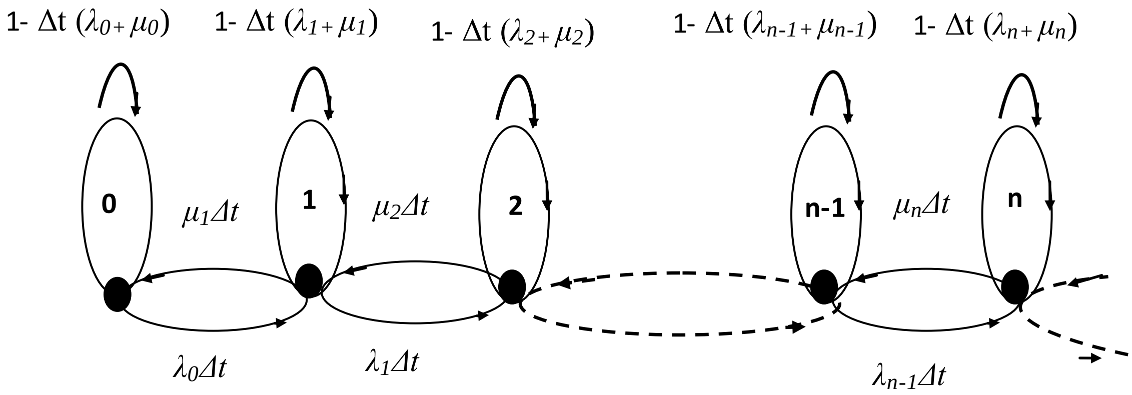


Figure II.4 : Processus de naissance et de mort

II.5.5.1 Matrice de transition

$$M = \begin{array}{c|cccccccc|} & E_0 & E_1 & E_2 & & E_{n-1} & E_n & E_{n+1} \\ E_0 & 1-\lambda_0\Delta t & \lambda_0\Delta t & 0 & \cdots & 0 & 0 & \cdots \\ E_1 & \mu_1\Delta t & 1-(\lambda_1+\mu_1)\Delta t & \lambda_1\Delta t & \cdots & 0 & 0 & \cdots \\ E_2 & 0 & \mu_2\Delta t & 1-(\lambda_2+\mu_2)\Delta t & \cdots & 0 & 0 & \cdots \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ E_{n-1} & 0 & 0 & 0 & \cdots & 1-(\lambda_{n-1}+\mu_{n-1})\Delta t & \lambda_{n-1}\Delta t & \cdots \\ E_n & 0 & 0 & 0 & \cdots & \mu_n\Delta t & 1-(\lambda_n+\mu_n)\Delta t & \cdots \\ E_{n+1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}$$

II.5.5.2 Equation régissant l'évolution du système

Soit : $\lambda_n \ \Delta t =$ {probabilité de passer de l'état n à la date t, à l'état $n+1$ à la date $t+\Delta t$} et $\mu_n \Delta t$ {probabilité de passer de l'état n à la date t, à l'état $n-1$ à la date $t+\Delta t$}, les $\lambda_n et$ μ_n étant positifs sauf pour μ_0 qui est nulle.

Soit encore $p_n(t)$ la probabilité pour le système d'être dans l'état n à la date t . On a selon les hypothèses précédentes :

En appliquant la relation $p(t+\Delta t) = p(t).M$ (où M est la matrice de transition), on trouve :

$$p_0(t+\Delta t) = (1-\lambda_0 \Delta t)p_0(t) + \mu_1 p_1(t)\,\Delta t$$

$$p_1(t+\Delta t) = \lambda_0 \Delta t p_0(t) + [1-(\lambda_1 + \mu_1)\,\Delta t]p_1(t) + \mu_2\, p_2(t)\Delta t$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$p_n(t+\Delta t) = \lambda_{n-1}p_{n-1}(t)\Delta t + [1-(\lambda_n + \mu_n)]p_n(t)\Delta t + \mu_{n+1}\, p_{n+1}(t)\Delta t$$

$$\text{avec } n \geq 1$$

D'où l'on déduit les équations différentielles :

$$p'_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t)$$

$$p'_n(t) = \lambda_{n-1}p_{n-1}(t) - (\lambda_n + \mu_n)\, p_n(t) + \mu_{n+1}p_{n+1}(t)$$

Si le nombre des états est limité, ce qui revient à dire que l'effectif total est limité à un nombre N, alors dans ce cas$\lambda_N = 0$ et la dernière équation différentielle devient :

$$p'_N(t) = \lambda_{N-1}p_{N-1}(t) - \mu_N\, p_N(t)$$

Dans le régime stationnaire c'est-à-dire quand $t \to \infty$ la distribution des probabilités d'états ne dépend plus du temps ; donc leur dérivées s'annulent. D'ou le système des équations différentielles précédent aura la configuration :

$$0 = \lambda_0 p_0 + \mu_1 p_1$$

$$0 = \lambda_{n-1}p_{n-1} - (\lambda_n + \mu_n)\, p_n + \mu_{n+1}p_{n+1} \qquad \forall\, n \geq 1.$$

De la première équation on tire $p_1 = (\lambda_0/\mu_1)p_0$.

pour $n = 1$, la deuxième équation donne :

$$0 = \lambda_0 p_0 - (\lambda_1 + \mu_1)p_1 + \mu_2 p_2$$

$$p_2 = \frac{\lambda_1\lambda_0}{\mu_2\mu_2} p_0$$

Et de proche en proche, on déduit :

$$p_n = \frac{\lambda_0\lambda_1 \dots \lambda_{n-1}}{\mu_1 \dots \mu_n} p_0$$

Pour déduire la valeur de p_0 on se base sur la propriété $\sum_{n=0}^{\infty} p_n = 1$. En remplaçant les p_n par leurs valeurs, on tire :

$$\sum_{n=0}^{\infty} p_n = 1 = p_0 + p_1 + p_2 + p_n + \dots$$

$$= p_0 + \frac{p_0\lambda_0}{\mu_1} + \frac{p_0\lambda_0\lambda_1}{\mu_0\mu_2} + \dots + \frac{p_0\lambda_{n-1} \dots \dots \lambda_1\lambda_0}{\mu_{n\dots}\mu_2\mu_1}$$

$$= p_0(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_0\mu_2} + \dots + \frac{\lambda_{n-1} \dots \dots \lambda_1\lambda_0}{\mu_{n\dots}\mu_2\mu_1}$$

Donc : $p_0 = \dfrac{1}{1 + \sum_{n=1}^{\infty}(\prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i})}$

II.6 Conclusion

Dans ce chapitre nous avons introduit les modèles mathématiques nécessaires pour l'étude du formalisme des files d'attente telles que : les chaîne de Markov, processus de Poisson, processus de naissance et de mort. Il est possible d'étudier ce formalisme de plusieurs manières, en le considérant soit comme une discipline abstraite de mathématique appliquées, soit comme un outil mathématique utile pour analyser ou évaluer le comportement d'un système d'attente.

Chapitre III :

Introduction aux files d’attente

III.1 Introduction

La Théorie des files d'attente est une technique de la Recherche opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et de déterminer ses caractéristiques pour aider les gestionnaires dans leurs prises de décisions.

Les files d'attente peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. On les rencontre dans les domaines d'activité les plus divers spécialement dans le trafic routier (carrefour à feux). L'étude mathématique des phénomènes d'attente constitue un champ d'application important des processus stochastiques.

On parle de phénomène d'attente chaque fois que certaines unités appelées “clients” se présentent d'une manière aléatoire à des “stations” afin de recevoir un service dont la durée est généralement aléatoire.

Il n'y a pas de théorie des files d'attente en tant que tel. La théorie des files d'attente est un formalisme mathématique qui permet de mener des analyses quantitatives à partir de la donnée des caractéristiques du flux d'arrivées et des temps de service.

L'objectif principal de ce chapitre est d'expliquer le phénomène d'attente, de présenter les notions de base concernant les systèmes de files d'attente et de définir les paramètres permettant de décrire les performances de tels systèmes.

III.2 Développement de la théorie des files d'attentes

Les origines du formalisme des files d'attente datent du début du XXème siècle et principalement des travaux de deux mathématiciens : le mathématicien danois A.K.Erlang avec ses travaux sur les réseaux téléphoniques et le russe A.A. Markov avec la création des modèles markoviens [20].

C'est en 1909 que les bases de ce formalisme sont jetées, grâce à l'article du mathématicien danois A.K. Erlang ''The theory of probabilities and telephone conversations". Les premiers résultats sont variés : Erlang observe le caractère poissonnier des arrivées des appels à un central téléphonique, et le caractère exponentiel des durées des appels; il réussit à calculer de manière relativement simple la probabilité d'avoir un appel rejeté. La notion d'équilibre stationnaire d'un système d'attente est introduite.

À partir des années 30, les travaux de plusieurs mathématiciens tels que Molina,Fry, Pollaczek aux États-Unis, Kolmogorov et Khintchine en Russie, Palm en Suède, ou Crommelin en France permettent à la théorie des files d'attente de se développer lentement. Ce sont ensuite les années 50 qui verront l'essor important de la théorie.

Les applications de ces travaux sont alors très pratiques, et concernent les disciplines de recherche opérationnelle et génie industriel. On peut citer les flux de trafic (véhicule, avion, personnes, télécommunications), l'ordonnancement, c'est-à-dire la planification, par exemple les patients dans les hôpitaux, les programmes d'un ordinateur, etc . . . ou encore le dimensionnement (banque, poste, réseaux, téléphone, ordinateur).

Dans les années 80, cette discipline devient beaucoup plus mathématique, et la littérature regorge d'articles décrivant des techniques ou des astuces mathématiques permettant de trouver des solutions exactes aux modèles.

Au cours de la décennie suivante, les chercheurs s'intéressent d'avantage à la création de modèles, et au calcul scientifique associé pour résoudre ces modèles. En effet, le développement de la puissance des ordinateurs permet maintenant d'obtenir des solutions approches des modèles suffisamment fiables pour être utilisées.

Actuellement ce sont les applications dans le domaine de l'analyse de performance des réseaux (téléphone mobile, Internet, multimédia, .) qui suscitent le plus de travaux.

III.3 Notations et définitions générales

Dans cette section nous commençons d'abord par la présentation des systèmes et réseaux de files d'attente avec leurs éléments caractéristiques puis, dans la deuxième partie, nous étudions les mesures de performances concernant les files d'attente.

Généralement une file d’attente est la donnée d’une (ou plusieurs) unités de services où arrivent des clients qui demandent une certaine durée d’utilisation de cette unité (le service demandé par les clients) .Quand les clients peuvent accéder à cette unité de service, ils patientent dans une file d’attente en attendant d’être servis .La file d’attente peut éventuellement n’accepter qu’un nombre fini des clients, dans ce cas les clients trouvant la file pleine à leur arrivée sont rejetés par le système. Un client peut être servi pendant une certaine période puis abandonné par le serveur. Le service résiduel d’un client est la durée du service qui reste à effectuer, quand celui-ci est nul, le client quitte la file d’attente.

La charge de la file d’attente est la somme de tous les services résiduels de tous les clients présents.

III.4 Modèle de Files d'attente

Tous les exemples de phénomènes d'attente ont des caractéristiques communes que l'on peut résumer ainsi : Des entités circulent dans un système et utilisent des ressources communes. Le système, les entités ou les ressources peuvent avoir un comportement imprévisible, c'est-à-dire dans le contexte d'une modélisation mathématique, aléatoire.
Une file d'attente se compose des éléments suivants :

1. La population.
2. Le nombre de serveurs.
3. Les tendances quant à l’arrivée et au service.
4. L’ordre de traitement des clients.
5. Une salle d'attente, c'est-à-dire un lieu ou les clients attendent quand aucun des serveurs n'est disponible pour les servir.

Tout système de file d’attente peut être représenté par le schéma suivant :

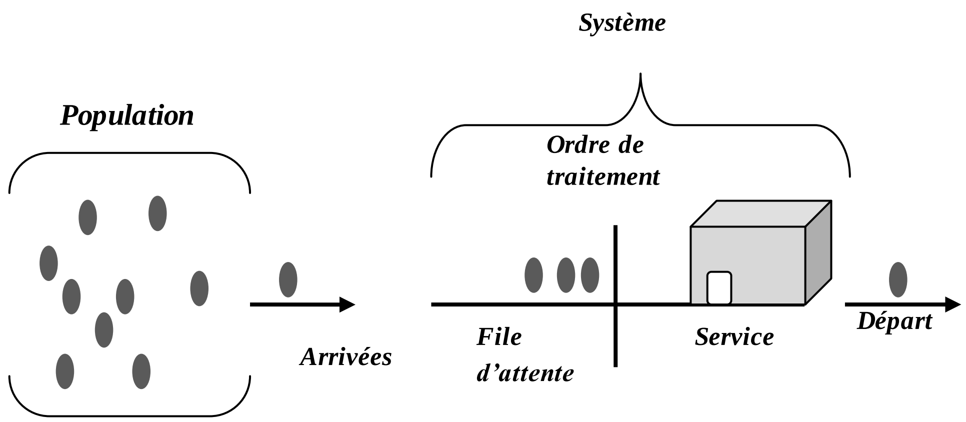


Figure III.1: Structure générale d'un système de file d'attente

III.4.1 La population

Dans la théorie des files d'attente, la population est la source de clients potentiels. Il y a deux situations possibles. Dans le premier cas, la population est infinie**,** c'est-à-dire que le nombre potentiel de clients est infiniment grand en tout temps. C'est le cas des clients des supermarchés, des banques, des restaurants, des cinémas, des centres d'appels, trafic urbain, etc. De plus, les clients proviennent de toutes les régions possibles.

Dans la deuxième situation, la population est finie, ce qui signifie que le nombre de clients potentiels est limité. Un bon exemple est le nombre de machines, d'avions, etc., en réparation dans le centre de maintenance d'une entreprise.

III.4.2 Le nombre de serveurs

La capacité de service dépend de la capacité de chaque serveur et du nombre de serveurs disponibles. Le terme « serveur » représente ici la ressource et, en général, on suppose qu'un serveur ne traite qu'un client à la fois. Les systèmes de files d'attente fonctionnent avec serveur unique ou serveurs multiples.

III.4.2.1 Systèmes de files d'attente avec un serveur unique

Ce type de systèmes comprend un seul serveur qui offre le service. Les exemples de systèmes de files d'attente avec serveur unique sont nombreux : les petits magasins avec une seule caisse, tels que les dépanneurs, certains cinémas, certains lave-autos et établissements de restauration rapide avec guichet unique.

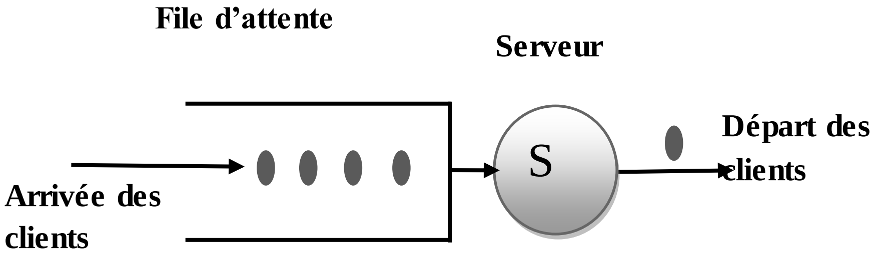


Figure III.2 : Système de file d'attente avec un serveur unique

III.4.2.2 Systèmes de files d'attente avec plusieurs serveurs

Ce type de systèmes comprend plusieurs serveurs qui fournissent le service. Les serveurs sont en parallèle ou en série.

a. Système de files d'attente avec plusieurs serveurs en parallèles

Chaque client ne requiert le service que d'un seul serveur et tous les serveurs sont capables de fournir ce service. Dans ce type de files d'attente, on a S serveurs en parallèle. Le client entrant au système n'est pas obligé de visiter tous les S serveurs. Si chaque serveur est doté d'une file, au moment de son arrivée le client choisit l'une d'entre elles ; bien sûr il doit choisir la file la moins longue ; ou bien il rejoint une position donné dans la file si elle est unique, puis il sera sélectionné de cette file selon la politique adoptée dans le système. Si n, le nombre de clients présentes dans le système est inférieur à *S*, ce qui revient à dire tant que les serveurs ne sont pas tous occupés, alors la file ne se constitue pas et tout client arrivant est immédiatement pris en charge par l'un des serveurs libres.

Si $n = S$ la file commence à se constituer et tout client qui arrive doit rejoindre la file.

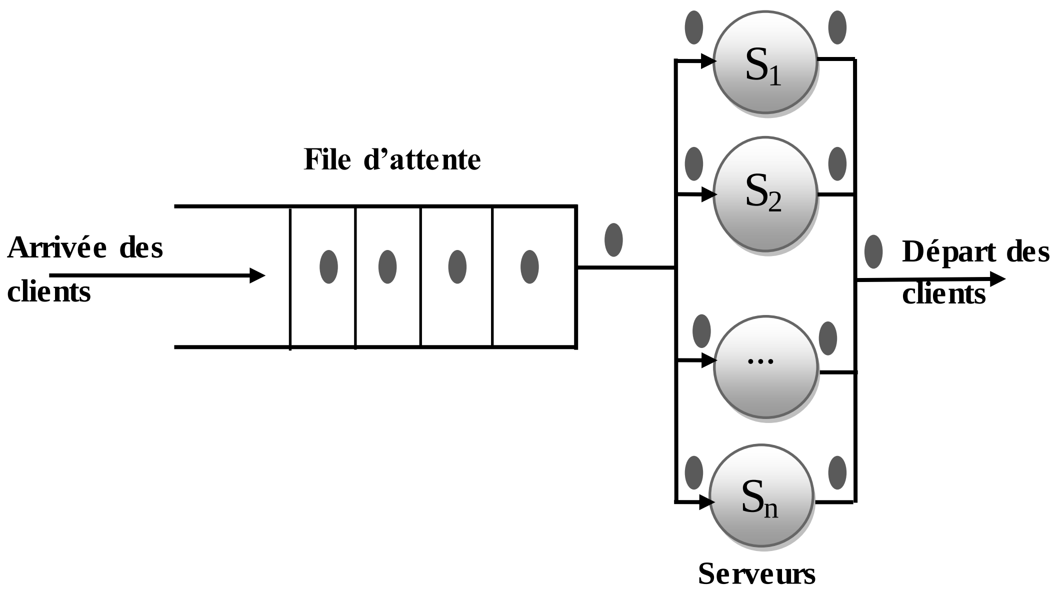


Figure III.3 : Système de file d'attente à serveurs parallèles

b. Système de files d'attente avec plusieurs serveurs en série

Dans ce type de files d'attente, on a S serveurs en série. Le client entrant au système doit visiter plusieurs serveurs successifs dans un ordre fixe pour recevoir le service. Ce type est appelé aussi système de files d'attente en cascade, on les rencontre dans une chaine de fabrication, dans la circulation des dossiers dans une administration…etc.

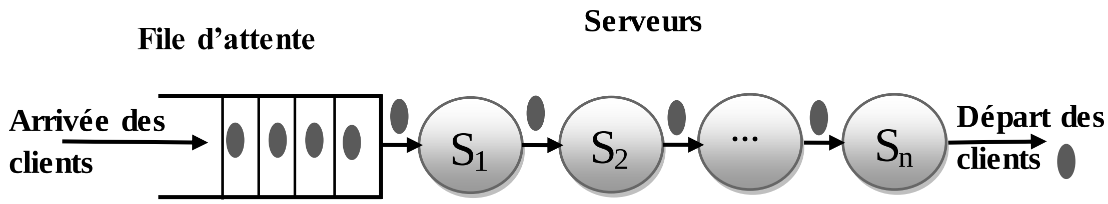


Figure III.4: Système de file d'attente à S serveurs en série

III.5 Le comportement d'une file d'attente

Cinq caractéristiques permettent de décrire le comportement d'une file d'attente :

- Le processus d'arrivée des clients;
- Le processus de service;

- La discipline de service de la file d'attente, c'est-à-dire la manière dont les clients sont choisis pour recevoir leur service;
- La capacité du système, c'est-à-dire le nombre total de clients pouvant se trouver dans le système à une instante donnée;
- Le nombre de serveurs;

1. Nomenclature de Kendall

Très souvent les arrivées et les durées de service de chaque client sont imprévisibles. Elles sont donc modélisées par des processus stochastiques, et sont alors caractérisées par les lois marginales du processus.

Une notation, dite notation de Kendall est communément utilisée pour décrire les cinq caractéristiques d'une file d'attente. Elle se présente sous la forme d'un symbole A/S/P/K/D, où chacune des lettres désigne une caractéristique de la file, comme précisée dans le tableau ci-dessous :

Symbole		***Désignation***
A	loi des inter-arrivées (durée entre deux arrivées successives)	
	M	distribution exponentielle (M=Markov)
	En	distribution d'Erlang à *n* phases
	D	distribution déterministe
	U	distribution uniforme
	G	distribution quelconque (G=General)
	GI	distribution quelconque, avec inter-arrivées deux à deux indépendantes
S	loi des durées de service des clients; Classification identique à celle des inter-arrivées;	
P	nombre de serveurs, $P \in \{1,2 \dots \dots \infty\}$	
K	capacité du système, c'est le nombre maximum de clients qui peuvent être présents simultanément dans le système, c'est-à-dire les clients en attente et les clients en service	
D	Discipline (ou politique) de service, précisant comment les clients sont servis	

Tableau III.1 : Nomenclature de Kandall

2. Processus des arrivées

Dans le formalisme des files d'attente, les arrivées des clients sont caractérisées par l'ensemble des instants ou dates d'arrivée de chaque client. Cette collection des dates d'arrivée s'appelle le processus des arrivées. Lorsque les dates d'arrivées sont imprévisibles, elles sont modélisées par des variables aléatoires, et le processus des arrivées est alors une collection de variables aléatoires, c'est-a-dire un processus stochastique.

Soit a_n la date d'arrivée du $n^{ième}$ client. Le processus des arrivées est alors la collection $\{a_n, n = 1,2,\ldots\}$de variables aléatoires. Lorsque les arrivées sont déterministes, ces variables aléatoires sont des constantes, le processus des arrivées devient une collection de réels positifs.

2.1 Inter-arrivée

On appelle inter-arrivée la différence entre deux dates d'arrivées successives. On note en général la $n^{ième}$ inter-arrivée τ_n qui s'exprime comme :

$$\tau_n = a_{n+1} - 1 \quad \text{(III. 1)}$$

Ainsi le processus des arrivées peut être complètement défini par la donnée de la suite des inter-arrivées$\{\tau_n, n = 1,2,\ldots\}$, si la date d'arrivée du premier client a_n est connue. En effet, en sommant l'équation (III. 1) pour $i = 1$ à $n - 1$, on obtient :

$$a_n = a_1 + \sum_{i=1}^{n-1} \tau i \quad \text{(III. 2)}$$

Souvent les arrivées des clients sont indépendantes. Ceci veut dire que les durées séparant deux arrivées successives sont des variables aléatoires indépendantes et de même loi .Les différentes formes du processus d’arrivées sont : arrivées régulières (les instants séparant deux arrivées successives sont constants et égaux à $1/\lambda$), arrivée Poissonnienne (les instants d’arrivées forment un processus de Poisson), arrivées suivant une loi d’Erlang d’ordre k (C’est un processus où les arrivées sont Poissonniennes mais seulement les clients d’ordre multiple de k sont considérés),etc..

2.2 Taux d'arrivée

On appelle taux d'arrivée des clients, ou intensité des arrivées, le nombre moyen de client qui arrivent dans le système par unité de temps.

2.3 Arrivées par groupe

Aux instants d'arrivée, plusieurs clients arrivent simultanément. On parle alors d'arrivées par groupe .La donnée de la loi caractérisant la taille du groupe est alors nécessaire pour déterminer le comportement du processus des arrivées.

2.4 Clients impatients

Les clients sont dits impatients lorsqu'ils quittent le système avant d'être servis. Cela peut arriver soit dès l'arrivée, s'ils jugent la file trop importante, on parle alors de découragement, soit après avoir attendu et c'est alors un abandon.

3 Processus de service

Un client entre dans une file d'attente pour utiliser des ressources qui sont modélisées par un serveur. Chaque client entrant dans une file d'attente va donc utiliser le serveur pendant une certaine durée qui dépend du client. Le processus de service pourra être d'une complexité extrême, dans la pluparts des systèmes de file d'attente la durée de service de chaque client est indépendante des autres, et qu'elles obéissent toutes à une même loi de distribution: on parle de variables indépendantes et identiquement distribuées (i.i.d.).

Soit σ_n le service demandé par le $n^{ième}$ client de la file ($n^{ième}$ arrivée). Les durées de service de tous les clients sont donc décrits par le processus de service $\{\delta_n , n = 1,2, \dots \}$, peut être déterministe ou aléatoire.

Comme pour les arrivées, le service peut être fourni par groupe, comme par exemple dans le cas d'un ordinateur avec traitement parallèle, ou encore un modèle de visite guidée ou d'embarquement dans un train.

Le processus de service peut dépendre du nombre de clients en attente, par exemple plus rapide si la file grossit, on parle alors de service dépendant de l'état du système.

Le processus de service peut être stationnaire ou non. Souvent les durées de service σ_n sont deux à deux indépendantes.

La date de départ ou date de sortie est la date de fin de service des clients. Elle est en général notée d_n, et peut s'exprimer en fonction de la date d'arrivée, le temps d'attente et la durée de service effectivement reçue. Si w_n dénote le temps d'attente du $n^{ième}$ client et si le service reçu est le service demandé, on a pour chaque client :

$$d_n = a_n + \delta_n + w_n \quad (III.3)$$

4 Discipline de service

La discipline de service détermine l'ordre dans lequel les clients sont rangés dans la file et y sont retirés pour recevoir un service. Les disciplines les plus courantes sont :

- **FIFO (**first in, first out) ou FCFS (first come first served) ou PAPS (premier arrivé, premier servi) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que dans la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui. En français, le terme PAPS comporte une ambiguïté, puisqu'il ne peut différencier une file "premier arrivé, premier servi" d'une file "premier arrivé, premier sorti".

- **LIFO** (last in, first out) ou LCFS (last come, first served) ou DAPS (dernier arrivé, premier servi). Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). A nouveau, les disciplines LIFO et LCFS ne sont pas équivalentes que pour une file mono-serveur.

- **RANDOM** (aléatoire).Le prochain client qui sera servi est choisi aléatoirement dans la file d'attente.

- **Round-Robin** (cyclique). Tous les clients de la file d'attente entrent en service à tour de rôle, effectuant un quantum Q de leur temps de service et sont replacés dans la file, jusqu' à ce que leur service soit totalement accompli. Cette discipline de service a été introduite afin de modéliser des systèmes informatiques ;

- **PS (Processor Sharing).** C'est le cas limite de la distribution Round-Robin lorsque le quantum de temps Q tend vers 0. Tous les clients sont servis en même temps, mais avec une vitesse inversement proportionnelle au nombre de clients simultanément présents. Si le taux du serveur est égal à μ et qu``a un instant donné il y a n clients à la station, tous les clients sont donc servis simultanément avec un taux μ/n

- **Priorité:** Il existe plusieurs classes de clients de priorité différente; le client avec la priorité la plus haute est servi en premier, et la discipline de service est FIFO au sein d'une même classe;

Pour la discipline ''Priorité", on distingue deux variantes, selon que le client en service est interrompu lorsqu'un client plus prioritaire arrive ou non. La discipline est une discipline de priorité non préemptive si le client en service n'est pas interrompu lorsqu'un client plus prioritaire entre dans le système. Dans le cas contraire, le client en service est interrompu lorsque qu'un client plus prioritaire arrive, afin que ce dernier puisse commencer son service aussitôt. Une fois le service du client prioritaire terminé, le client interrompu reprend son service, et on distingue deux politiques :

- **Priorité préemptive avec recommencement:** le client interrompu reprend son service au début;
- **Priorité préemptive avec continuation:** le client interrompu reprend son service là où il avait été interrompu.

5. Capacité de la file

La capacité de la file à accueillir des clients en attente de service peut être finie ou infinie.

Soit K la capacité de la file (incluant le ou les clients en service). Une file à capacité illimitée vérifie $K = +\infty$. Lorsque la capacité de la file est limitée et qu'un client arrive alors que cette dernière est pleine, le client est perdu.

6. Notion de classes de clients

Une file d'attente peut être parcourue par différentes classes de clients. Ces différentes classes se distingueront par :

- des processus d'arrivée différents ;
- des temps de service différents ;
- un ordonnancement dans la file d'attente en fonction de leur classe.

Pour définir une file multi-classe, il faut définir pour chaque classe de clients le processus d'arrivée et la distribution du temps de service associe. Il faut également préciser comment les clients des différentes classes s'ordonnent dans la file d'attente.

III.6 Les réseaux de files d'attente

Pour arriver à modéliser des systèmes beaucoup plus complexes, une file d'attente simple n'est pas suffisante, il faut faire appel à des réseaux de files d'attente.

Définition : Un réseau de files d'attentes est la donné de plusieurs files d'attente entre lesquels circulent des flots de clients. Le réseau peut être ouvert, fermé, ou mixte.

III.6.1 Les réseaux ouverts

Dans un réseau de files d'attente ouvert, les clients arrivent de l'extérieur, circulent dans le réseau à travers les différentes stations, puis quittent le réseau. Le nombre de clients pouvant se trouver à un instant donnée dans un réseau ouvert n'est donc pas limité. Afin de spécifier complètement un réseau ouvert, il faut bien sûr caractériser chaque station, mais également le processus d'arrivée des clients et le routage (cheminement) des clients dans le réseau.

a- Processus d'arrivée

Le processus d'arrivée des clients dans le réseau est décrit, comme pour une file simple, à l'aide d'un processus de renouvellement (il est donc caractérisé par la distribution du temps d'inter-arrivée). Si l'arrivée des clients suit un processus de Poisson, les inter-arrivées sont exponentielles et sont caractérisées par un unique paramètre : le taux d'arrivée λ. Dans le cas d'un processus d'arrivée non *Poissonnien*, ce paramètre reste intéressant, puisqu'il indique le nombre moyen de clients qui arrivent dans le système par unité de temps, mais devient insuffisant pour caractériser parfaitement l'arrivée des clients.

b- Routage des clients

Lorsqu'un client termine son service à une station, il faut préciser où ce client va se rendre : soit à une autre station, soit à l'extérieur (le client quitte alors le réseau).

Il existe cependant d'autres types de routages :

- le routage vers la file la plus courte (routage dynamique) : un client quittant une station choisira, parmi toutes les destinations possibles, la station qui comporte le moins de clients

- le routage cyclique (routage déterministe) : les clients quittant une station choisiront à tour de rôle chacune des stations parmi toutes les destinations possibles.

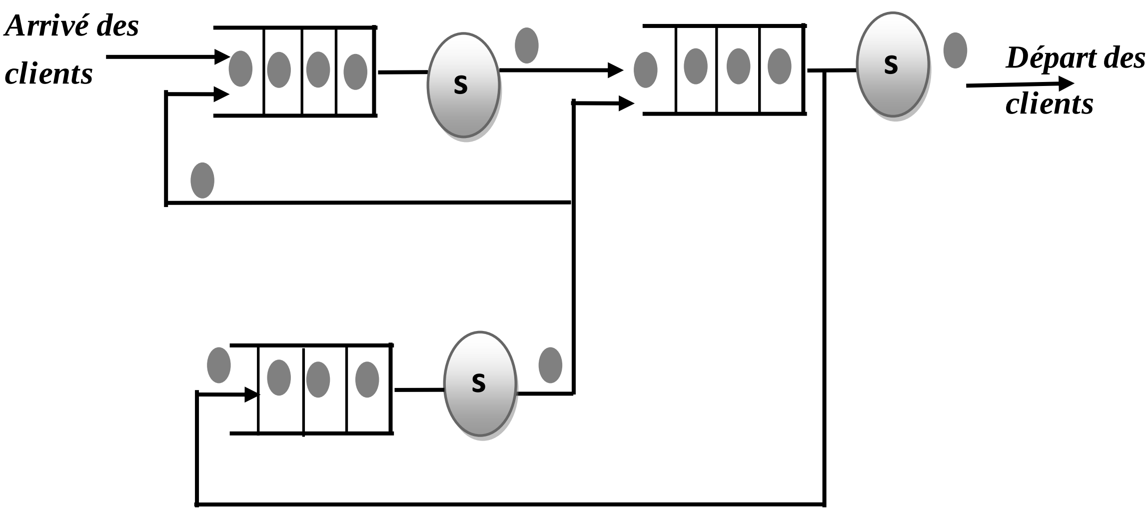


Figure III.5 : Un réseau de files d'attente ouvert

III.6.2 Les réseaux fermés

Dans un réseau de files d'attente fermé, les clients sont en nombre constant. Soit N le nombre total de clients du système. Il n'y a donc pas d'arrivée ni de départ de clients. La spécification d'un réseau fermé se réduit donc à celle des différentes stations et à celle du routage des clients.

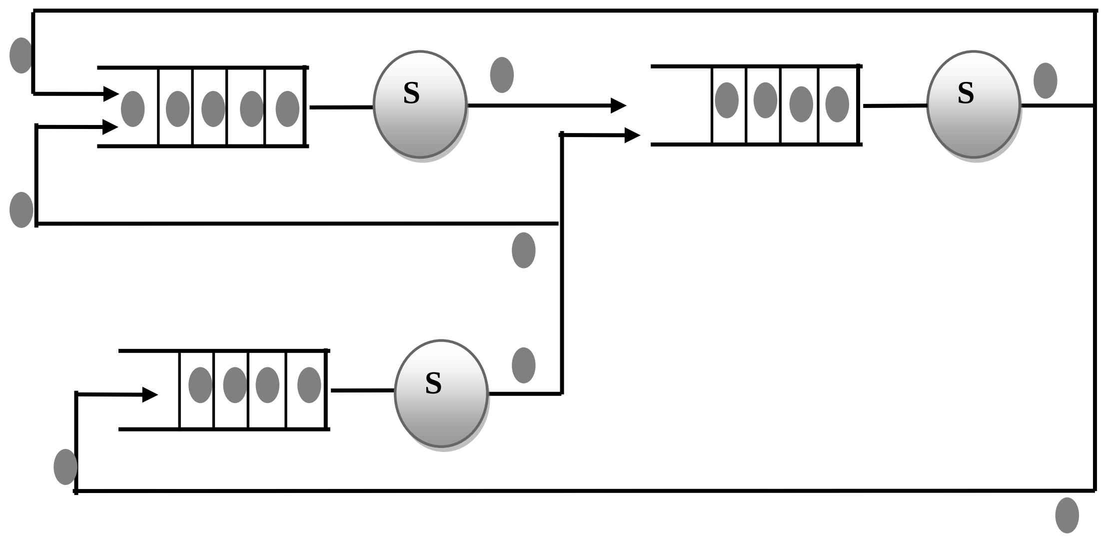


Figure III.6: Un réseau de files d'attente fermé

III.6.3 Les réseaux multi-classes

Comme pour les files simples, les réseaux de files d'attente peuvent être parcourus par différentes classes de clients. Soit R le nombre de classes de clients. Ces différentes classes se distingueront par:

- des processus d'arrivée différent (si le réseau est ouvert)
- des comportements différents à chaque station (service et discipline de service)
- des routages différents dans le réseau.

Il faut donc caractériser pour chaque classe r :

- pour un réseau ouvert, le processus d'arrivée (pour un processus d'arrivée Poissonnienne, il suffit alors de donner le taux d'arrivée des clients de classe r) ;
- pour un réseau fermé, le nombre total N_r de clients de classe r ;
- le routage de clients.

La notion de réseaux multi-classes nous permet d'introduire la notion de réseau mixte qui est un réseau ouvert vis à vis de certaines classes et fermé vis à vis des autres classes.

III.6.4 Les réseaux de files d'attente à capacité limitée

Les différentes stations du réseau peuvent avoir des capacités limitées. Lorsqu'une file est pleine, plus aucun client ne peut y entrer. Cela introduit des blocages dans les autres stations amont et éventuellement des pertes de clients à l'entrée du système (si celui-ci est ouvert).

On distingue principalement deux types de blocage : le blocage avant service et le blocage après service.

- Dans un blocage avant service (ou blocage de type "réseau de communication"), un client voulant commencer son service à une station donnée doit tout d'abord s'assurer qu'il y a une place libre dans la station de destination. Si c'est le cas, son service commence. Dans le cas contraire, le serveur de la station est bloqué et le client doit attendre la libération d'une place en aval avant de commencer son service.

- Dans un mécanisme de blocage après service (ou blocage de type "système de production"), un client commence sans attendre son service dés l'instant où le serveur est disponible. Ce n'est qu'à la fin de son service qu'un blocage peut survenir. Si la station de destination est pleine, le client reste au niveau du serveur qui se trouve alors bloqué, jusqu'à ce qu'une place se libère en aval.

III.6.5 Les réseaux de files d'attente ouverts à contrainte de population

Certains réseaux de files d'attente, bien qu'étant des modèles ouverts, peuvent être soumis à une limite supérieure sur le nombre total de clients pouvant s'y trouver simultanément. Cette "contrainte de population" implique que le réseau n'est ni réellement un modèle ouvert, puisque le nombre de clients qui peuvent s'y trouver est limité, ni réellement un réseau fermé, puisque le nombre total de clients dans le système n'est pas constant. On parlera de "modèle à contrainte de population". Lorsqu'un client arrive dans le réseau alors que celui-ci est plein (la contrainte de population est atteinte), deux cas peuvent être envisagés. Soit le client est "rejeté", ce qui rejoint le modèle de la section précédente, soit le client est “mémorisé” et se place en attente dans une file externe (généralement FIFO).

Un système ouvert à contrainte de population est souvent modélisé à l’aide d’un formalisme de type “sémaphore”. Une file de “jetons” contenant initialement N jetons est alors associée à la file externe des clients. Lorsqu’un client arrive alors qu’il reste un jeton de libre, il prend le jeton et entre instantanément dans le système. Il conserve alors le jeton pendant tout son séjour dans le système et le libère dés qu’il quitte le système. Le jeton revient alors instantanément dans la file des jetons et devient à nouveau disponible pour un autre client. Lorsqu’un client arrive alors qu’il n’y a aucun jeton de libre, il se place dans la file externe (des clients) en attente de libération d’un jeton. Le nombre initial N de jetons impose donc une limite supérieure sur le nombre total de clients pouvant se trouver simultanément dans le système.

III.7 Etude analytique d’un système de file d’attente M/M/1

Après la présentation des différents systèmes de files d’attente, nous nous intéressons maintenant à l’analyse d’un système de file d’attente M/M/1[1] dont le but de déterminer les paramètres permettant de décrire les performances de tels systèmes.

L'efficacité d'une file d'attente est déterminée par un certain nombre de mesures de performance qui sont calculées ou évaluées à partir de ses caractéristiques physiques.
De manière générale on peut identifier trois familles de mesures de performance intéressantes à calculer lors de l'analyse d'une file d'attente :

- Une mesure du temps d'attente subit par un client quelconque;
- Une indication sur la manière dont les clients peuvent s'accumuler dans le système;
- Une mesure de la durée de repos ou d'activité des serveurs.

Comme les caractéristiques de base des files sont souvent aléatoires, ces éléments de mesure de performance sont également aléatoires, et leur caractérisation nécessite la détermination des lois associées ou au moins des valeurs moyennes.

[1] M/M/1 : file d’attente markovien avec un serveur unique

III.7.1 Structure générale d'un système M/M/1

Le système de file d'attente M/M/1 est un système où l'arrivée des clients est Poissonienne de taux λ (nombre moyen de clients arrivant pendant une unité de temps) et la durée du service est exponentielle de taux µ (nombre moyen de clients servis pendant une unité de temps, donc $\frac{1}{\mu}$ est le temps moyen de service d'un client). Une file M/M/1 est donc une file avec un processus de Markov en entrée et en sortie, un seul serveur, une discipline de service premier arrivé, premier servi, une capacité infinie et un nombre infini de clients qui peuvent entrer dans cette file. L'exemple le plus courant est l'attente des voitures devant un carrefour à feux.

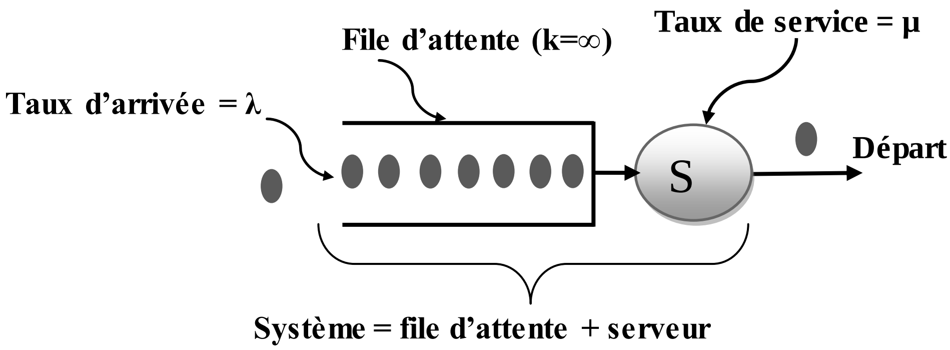


Figure III.7 : *Structure générale d'un système de file d'attente M/M/1*

III.7.2 Etude analytique du système

Le system M/M/1 se rapporte facilement à un processus de naissance et de mort ou $X(t)$ représente le nombre de clients dans le systeme, chaque arrivée est considérée comme une naissance et chaque fin de service est considérée comme une mort.

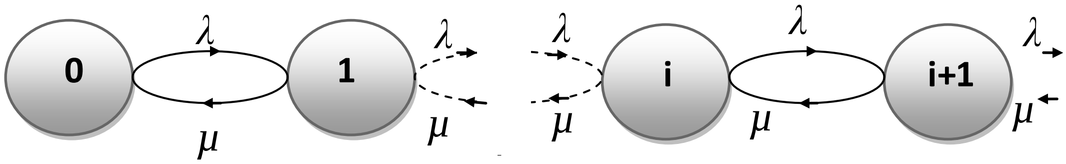


Figure III.8:Graphe d'état du système M/M/1

Le nombre de clients dans le système à un instant quelconque est un nombre entier, et l'espace d'état du processus $\{X(t),\ t \geq 0\}$ donc $\xi = N$.

Soient $0 \leq t_0 < t_1 < \ldots < t_{n+1}$ des instants déterministes quelconques et n un entier arbitraire. Nous devons montrer la propriété de Markov :

$P\ \{X_{t+1} = j / X_0 = a, X_1 = b, \ldots, X_t = i\} = P\ \{X_{t+1} = j / X_t = i\}$ Le nombre de clients dans le système à l'instant t_{n+1}, $X\ (t_{n+1})$ dépend :

- du nombre de clients dans le système à l'instant t_n, $X\ (t_n)$,
- du nombre d'arrivées dans l'intervalle : $]\ t_n, t_{n+1}]$,
- et du nombre de clients quittant le système, c'est-à-dire terminant leur service, dans l'intervalle : $]\ t_n, t_{n+1}]$,

Le nombre d'arrivées sur des intervalles de temps disjoints étant indépendant (processus de Poisson) et les durées de services étant deux à deux indépendantes et indépendantes des arrivées, sachant $\{X(t_n) = i\}$ la dépendance de $X(t_{n+1})$ avec le passé de la chaîne ne peut se faire qu'au travers de la durée de service résiduelle du client éventuellement en service à t_n et de la durée avant la prochaine arrivée après t_n.

Le processus des arrivées étant un processus de Poisson, les durées inter-arrivées sont indépendantes et identiquement distribuées, de loi exponentielle. D'après la propriété d'absence de mémoire de cette loi, la durée résiduelle avant l'arrivée suivant l'instant t_n suit elle-même une loi exponentielle et est indépendante du passé. Il en est de même pour la durée de service résiduelle du client éventuellement en service à t_n, et on déduit aisément la propriété de Markov.

Les hypothèses d'indépendance que cela soit entre les processus des arrivées et de services, ou entre les inter-arrivées ou les durées de services sont primordiales, tout comme la propriété d'absence de mémoire de ces durées.

III.7.2.1 Détermination du régime permanent / Conditions de stabilité

Pour pouvoir étudier ce système entre les instants t et $t + \Delta t$, considérons l'un de ses états, soit l'état n. Cet état désigne qu'il existe n clients dans le système à l'instant $t + \Delta t$.

Ce cas a lieu dans les situations suivantes :

- à l'instant t il y a eu $(n-1)$ clients dans le système et une arrivée a eu lieu pendant Δt, mais aucun service n'a été achevé pendant cet intervalle. Donc à $t+\Delta t$ on a un effectif de n. La probabilité correspondante sera :

$$p_n\ (t+\Delta t)\ =\ p_{n-1}(t)[\lambda\Delta t(1-\mu\Delta t)]\ \ (III.12).$$

-à l'instant *t* il ya eu n clients dans le système, pendant l'intervalle Δt aucune entrée et aucune achèvement de service; donc aucune sortie n'a eu lieu. L'effectif à $t+\Delta t$ sera donc n et la probabilité correspondante sera :

$$p_n\ (t+\Delta t)\ =\ p_n(t)(1-\lambda\Delta t)(1-\mu\Delta t)$$

- à l'instant t il ya eu $(n+1)$ clients dans le système, un service a été achevé pendant Δt, et aucune entrée n'a été enregistrée pendant cet instant. Donc l'effectif à $t+\Delta t$ sera n et la probabilité correspondante :

$$p_n\ (t+\Delta t)=\ p_{n+1}(t)(1-\lambda\Delta t)\mu\Delta t.\ (III.13).$$

Donc :

$$p_n\ (t+\Delta t)=\ p_{n-1}(t)[\lambda\Delta t(1-\mu\Delta t)]$$
$$+\ p_n(t)(1-\lambda\Delta t)(1-\mu\Delta t)+p_{n+1}(t)(1-\lambda\Delta t)\mu\Delta t\ \ (III.14).$$

Le développement de cette expression nous donne :

$$p_n\ (t+\Delta t)=\ p_{n-1}(t)\lambda\Delta t\, p_n(t)[(1-(\lambda+\mu)\Delta t)\ +\ p_{n+1}\mu\Delta t+0(\Delta t^2)\qquad (III.15).$$

Où $0(\Delta t^2)$ est un infiniment petit en(Δt^2) .
Quand $\Delta t \rightarrow 0$, l'équation différentielle correspondante sera :

$\frac{Pn(t+\Delta t)-Pn(t)}{\Delta t}$ tend vers P_n' donc :

$$P_n'(t)=\ p_{n-1}(t)\lambda\ -\ p_n(t)\ (\lambda+\mu)\ +\ p_{n+1}(t)\ \ \mu\ \ (III.16)$$

À l'état stationnaire les différentielles s'annulent et les probabilités ne dépendent plus de temps. D'ou l'équation précédente s'écrira alors sous la forme :

$$0=\ \lambda\, p_{n-1}(t)-\ (\lambda+\mu)\ p_n\ \ +\ \mu\ \ p_{n+1}\qquad \forall n\geq 1\qquad (III.17)$$

Pour n=1, c'est-à-dire quand on n'a aucun client dans le système à $t+\Delta t$ le raisonnement sera comme suit :

-à l'instant t il n'ya a eu aucun client dans le système et aucune arrivée n'a eu lieu pendant Δt. La probabilité correspondante sera :

$$p_0(t)(1-\lambda\Delta t)$$

-à l'instant t il y a eu un seul client en service et il a achevé son service pendant Δt, mais aucune arrivée n'a eu lieu pendant cet intervalle. D'où la probabilité de réalisation :

$$p_1(t)\mu\Delta t(1-\lambda\Delta t).$$

Donc : $p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\mu\Delta t(1-\lambda\Delta t)$.

Après avoir développé l'expression et éliminé les termes en $(\Delta t)^2$on obtient :

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\mu\Delta t.$$

Alors : $\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}$ tend vers $P_0'(t)$ donc :

$$P_0'(t) = \lambda p_0(t) + \mu p_1(t). \quad (III.18).$$

A l'état stationnaire on peut écrire :

$$0 = \lambda p_0 + \mu p_1 \qquad \Rightarrow p_1 = \frac{\lambda}{\mu} p_0 \qquad (III.19).$$

En portant la valeur de P_1dans l'expression de P_2 qui pourrait être déduite de l'équation (III. 19) pour $n = 1$, on trouve :

$$0 = \lambda p_0 - (\lambda_n + \mu_n)\frac{\lambda}{\mu} p_0 + \mu p_2$$

$$\Rightarrow p_2 = \left(\frac{\lambda}{\mu}\right)^2 p_0$$

De proche en proche, on déduit :

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0 \text{ Avec } n \geq 1 \quad (III.20)$$

Considérons $q = \frac{\lambda}{\mu}$ (taux ou intensité de trafic) avec $q < 1$ est une condition essentielle pour la stabilité du système et elle est dit condition d'ergodicité du système. Avec l'équation de normalisation, on peut écrire :

$$\sum_{n=0}^{\infty} p_n = 1 \text{ , soit } \sum_{n=0}^{\infty} (\frac{\lambda}{\mu})^n \, p_0 = 1$$

Alors $\sum_{n=0}^{\infty} q^n \, p_0 = 1$

Impose alors que la série$(\sum_{n=0}^{\infty} q^n) = 1$ converge pour que les équations d'équilibre admettent une solution strictement positive. La série converge si et seulement si

$q < 1$, soit $\lambda < \mu$, et sa somme est alors $\frac{1}{1-q}$ ce qui donne :

$p_0 = 1 - q$. Les probabilités p_n valent alors :

$$p_n = (1 - q)q^n \quad \forall n \geq 1 \quad (III.21)$$

Si $q < 1$, soit $\lambda < \mu$, la file d'attente est stable, il existe un régime dit stationnaire ou permanent : le nombre de clients dans le système à l'instant $t, X(t)$ converge en loi lorsque $t \rightarrow \infty$ vers une variable aléatoire discrète N de probabilités élémentaires $p_n(1 - q)q^n$

et on a pour $j = 0,1,2,3 \ldots.$

$$\lim_{\rightarrow\infty} P \; [X(t) = j / X(0) = i] = \lim_{\rightarrow\infty} P \; (X(t) = j)$$
$$= p_n = (1 - q)q^n$$

Si $q \geq 1$, la file d'attente est instable, il n'existe pas de régime permanent ou stationnaire. La file grossit indéfiniment.

Remarque : Il est important de remarquer que le calcul des probabilités en régime permanent ne dépend pas de la discipline de service choisie. Les probabilités p_n trouvées restent valables pour toute discipline de service pour laquelle la durée de service effectivement réalisée est la durée de service demandée par le client.

III.7.2.2 Mesure de performance d'un système d'attente M/M/1

Afin de faciliter la compréhension des paramètres de performance, le tableau suivant présente les symboles et la terminologie utilisés pour l'étude des systèmes de file d'attente.

symbole	Signification
λ	Taux d'arrivée des clients
μ	Taux de service
$1/\mu$	Temps de service
q	Taux d'utilisation du système
$\tilde{n}_f$	Nombre moyen de clients qui attendent d'être servis
$\tilde{n}_s$	Nombre moyen de clients dans le système (clients qui attendent et clients qui sont en train d'être servis)
T_f	Temps moyen d'attente en file
T_s	Temps moyen d'attente dans le système (temps d'attente en file, plus le temps de service)
P_0	Probabilité qu'il y ait zéro unité (client) dans le système
P_n	Probabilité qu'il y ait n unités (clients) dans le système

Tableau III.2 : les paramètres de performance d'un système de file d'attente

Avant la détermination des paramètres de performance du système, nous présentons d'abord la loi de Little qui joue un rôle très important dans l'étude des systèmes des files d'attente.

-La loi de Little

La loi de Little est une relation très générale qui s'applique à une grande classe de systèmes. Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'inter-arrivées, temps de service,...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable. Le débit du système est alors indifféremment, soit le débit d'entrée, soit le débit de sortie : $d_s = d_e = d$. La loi de Little s'exprime telle que dans la propriété suivante :

Le nombre moyen de clients dans un système est égal au produit du taux moyen d'arrivées par le temps moyen passé par chaque client dans le système.

$$N = \lambda.T$$

-Démonstration : Pour simplifier la démonstration, supposons que le système soit vide à l'instant initial, soit $N(0) = 0$, et que la discipline de service soit FIFO.

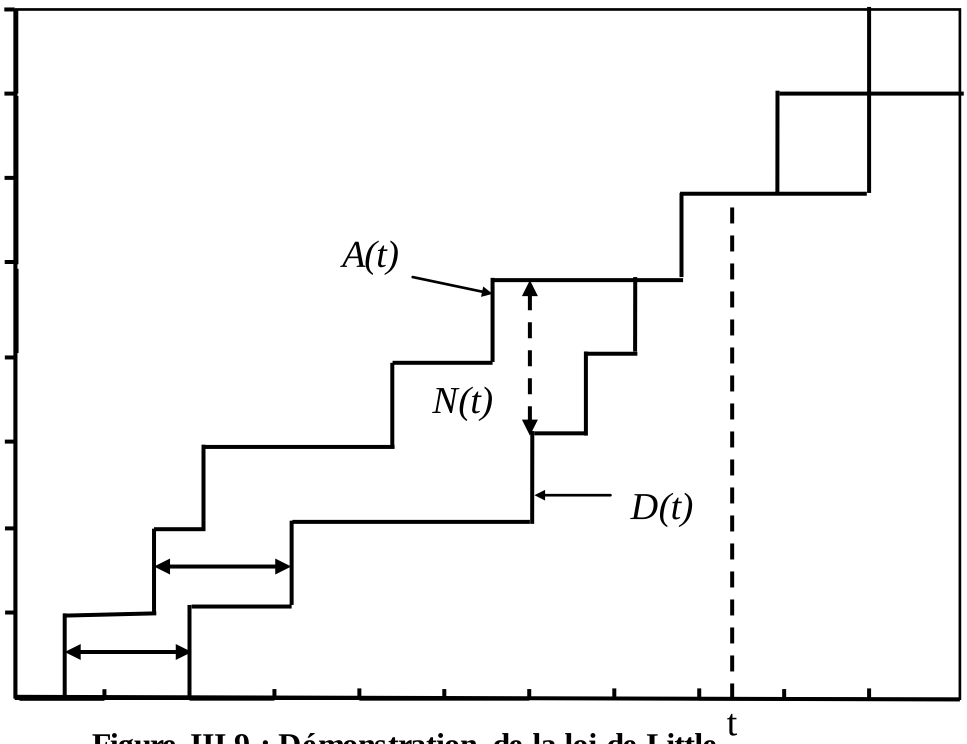


Figure III.9 : Démonstration de la loi de Little

La figure III.9 représente les graphes des fonctions $A(t)$ et $D(t)$ en fonction du temps t. Les fonctions $A(t)$ et $D(t)$ représentent respectivement le nombre total d'arrivées et de départs dans l'intervalle $[0;\ t]$, et sont donc des fonctions en escalier croissantes, d'incrément unitaire. Les sauts correspondent à une arrivée pour $A(t)$ ou à un départ pour $D(t)$. A tout instant t, la différence $A(t) - D(t)$ représente le nombre de clients dans le système à cet instant. L'aire comprise entre les deux courbes $A(t)$ et $D(t)$ sur le graphe de la figure III.9 est l'intégrale de la fonction $A(t) - D(t) = N(t)$, et s'exprime donc comme :$\int_0^t N(s)ds$

Pour tout instant t tel que $N(t) = 0$, (c'est-à-dire $A(t) = D(t)$), cette aire s'exprime également comme la somme des aires de chacun des rectangles la constituant. La différence entre la date de départ d_i et l'instant d'arrivée a_i du client n'est rien d'autre que le temps passée dans le système par le client (noi_j), soit Ti. On peut donc écrire :

$$\int_0^t N(s)ds = \sum_{i=1}^{A(t)} T\,i \quad \text{pour tout t tel que } N(t) = 0$$

Il reste maintenant à prendre la limite quand $t \to \infty$ dans l'expression ci-dessous pour obtenir :

$$N_t = \frac{1}{t}\int_0^t N(s)ds = \frac{A(t)}{t}\frac{1}{A(t)}\sum_{i=1}^{A(t)} T_t = \lambda_t T_t$$

qui est l'expression désirée. Il est important de noter que le passage à la limite ci-dessus n'est possible que si l'on sait que le système se vide (soit $N(t) = 0$) infiniment souvent et pour des temps t arbitrairement grand. Ceci est en fait vrai (et admis ici) pour une file d'attente à l'équilibre.

Remarquons qu'en l'absence de cette propriété, il est possible, quelque soit l'instant t, de majorer l'aire $\int_0^t N(s)\,ds$ par $\sum_{i=1}^{A(t)}$ Ti et de la minorer par $\sum_{i=1}^{D(t)}$ Ti, et on en déduit alors pour tout $t > 0$,

$$\frac{D(t)}{t}\frac{1}{D(t)}\sum_{i=1}^{D(t)} T_i \leq \frac{1}{t}\int_0^t N(s)ds \leq \frac{A(t)}{t}\frac{1}{A(t)}\sum_{i=1}^{A(t)} T_i$$

Soit:

$$\frac{D(t)}{t}\frac{1}{D(t)}\sum_{i=1}^{D(t)} T_i \leq N_t \leq \lambda_t T_t$$

En admettant que pour une file d'attente à l'équilibre $D(t) \to \infty$ tend vers le taux d'arrivée moyen λ lorsque $t \to \infty$ et

$$\lim_{t\to\infty} \frac{D(t)}{t}\frac{1}{D(t)}\sum_{i=1}^{D(t)} T_i$$

$$= \lim_{t\to\infty} \frac{A(t)}{t} \frac{1}{A(t)} \sum_{i=1}^{A(t)} T = T$$

On obtient $N = \lambda\, T$ en prenant la limite lorsque $t \rightarrow 1$ dans l'expression ci-dessus. Signalons enfin que cette preuve graphique peut être facilement adaptée pour démontrer la loi de Little sans l'hypothèse d'une discipline de service FIFO.

L'importance de la loi de Little tient dans sa généralité. Elle est en fait valable pour quasiment tout système ''stable", c'est-a-dire tout système qui possède un régime permanent.

-Nombre moyen de clients dans le système

Le nombre n de clients dans le système est égal au nombre de clients en attente plus le nombre de clients en cours de service. La valeur moyenne de n n'est autre que l'espérance mathématique de l'état du système.

$$\tilde{n} = E(n) = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n.(P_0\, q^n)$$

$$= (1 - q) \sum_{n=0}^{\infty} nq^n$$

$$=(0q^0 + 1q^1 + 2q^2 + 3q^3 + \ldots) - (0q^1 + 1q^2 + 2q^3 + 3q^4 + \ldots)$$

$$= 0 + q + q^2 + q^3 + \ldots = \sum_0^{\infty} q^n - 1$$

Lorsque $q<1$, $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$, on a finalement :

$$\tilde{n} = \frac{q}{1-q} \qquad \text{(III.21).}$$

- Nombre moyen de clients dans la file

Dans un système à un seul serveur : $n_f = n$ si $n = 0$ et $n_f = n - 1$ si $n \geq 1$

$$\tilde{n}_f = \sum_{n=1}^{\infty} n_f\, .P_n$$

$$= \sum_{n=1}^{\infty} (n-1)\, P_n + 0P_0$$

$$= \sum_{n=1}^{\infty} (n-1)(1-q)q^n$$

$$=(1-q)(0q+q^2+q^3+q^4+\dots)$$

$$=(0q^1+1q^2+2q^3+3q^4+\dots)-(0q^2+q^3+2q^4+\dots)$$

$$=q^2+q^3+q^4+\dots$$

$$=\sum_0^\infty q^n - q^0 - q^1$$

$$=\sum_0^\infty q^n - (1+q)$$

$$=\frac{1}{1-q} - (1+q)=\frac{1-(1-q^2)}{1-q}$$

Ou encore, en portant dans l'équation (III.21): $\tilde{n}_f = \tilde{n}.q$

$$\text{alors : } \tilde{n}_f = \frac{q^2}{1-q} \qquad (III.22).$$

Les caractéristiques qui concernent le temps de séjour et le temps d'attente peuvent être obtenues d'une façon purement intuitive en se basant sur les hypothèses du système.

- Temps moyen de séjour dans le système

En chaque unité de temps il arrive en moyenne λ clients et part en moyenne λ clients (régime permanent), l'intervalle moyen de temps entre deux arrivées est donc $1/\lambda$. Au moment de l'arrivée d'un nouveau client il y aura n clients dans le système, et avant son départ il y a en moyenne l'arrivée de n clients. Le temps que passe un client dans le système T_S est égal donc au temps nécessaire à l'arrivée de n clients, donc :

$$T_s = \frac{1}{\lambda}\tilde{n}$$

On portant dans (III.21)

$$T_s = \left(\frac{q}{1-q}\right)\frac{1}{\lambda} = \left(\frac{\lambda/\mu}{1-\lambda/\mu}\right)\frac{1}{\lambda}$$

$$T_s = \frac{1}{\mu-\lambda} = \frac{q}{\lambda(1-q)} \qquad (III.23).$$

- Temps moyen d'attente

En moyenne, le temps que passe un client en attente est égal au temps nécessaire à l'arrivée de $\tilde{n}_f$ clients :

$$T_f = \frac{1}{\lambda}\ \tilde{n}_f$$

$$T_f = \left(\frac{q^2}{1-q}\right)\frac{1}{\lambda} = \left(\frac{(\frac{1}{\lambda})2}{1-\lambda/\mu}\right)\frac{1}{\lambda} = \frac{\lambda/\mu 2}{\frac{\mu-\lambda}{\mu}} = \left(\frac{1}{\mu-\lambda}\right)\frac{\lambda}{\mu}$$

$$T_f = q.\ T_s \quad \text{(III.24).}$$

D'une manière analogue on dira qu'un nouvel arrivant attendra que les $\tilde{n}$ clients qu'il trouvera dans le système soient servis, le temps de service pour un client étant $1/\mu$; le temps d'attente qu'il subira est donc : $Tf = \tilde{n}.(1/\mu)$.

D'ailleurs ce résultat peut être obtenu à partir de l'équation (III.18) de la manière suivante

$$T_f = \frac{1}{\lambda}\ T_s$$

$$T_f = \left(\frac{q^2}{1-q}\right)\frac{1}{\lambda} = \left(\frac{q}{1-q}\right)\frac{q}{\lambda}$$

Mais $\frac{q}{1-q} = \tilde{n}$, donc :

$$T_f = \frac{1}{\mu}\tilde{n} = \frac{q^2}{\lambda(1-q)} \quad \text{(III .25).}$$

Remarque

En régime permanent, le temps moyen de séjour est égal à la somme du temps moyen d'attente T_f et du temps moyen de service $\frac{1}{\mu}$ et on retrouve bien :

$$T_s = \frac{q^2}{\lambda(1-q)} + \frac{1}{\mu} \quad \text{(III.26).}$$

- Probabilité de dépasser un nombre k de clients dans le système

Une autre quantité importante à calculer est la probabilité de dépasser un nombre k de clients dans le système. Cette quantité sert par exemple à prévoir les moyens à investir pour accueillir les usagers en matière d'espace, d'aération, de chaises, etc. Elle sert également à estimer combien de fois (en pourcentage) l'investissement consenti en terme de temps de service surtout mais également en terme de condition d'accueil sera insuffisant.

$$P(n \geq k) = \sum_{i=1}^{k} Pi$$

$$= \sum_{i=1}^{k} (1-q)q^{i} = \sum_{i=1}^{k} (q^{i}) - \sum_{i=1}^{k} (q^{i+1})$$

$$P(n \geq k) = q^{k} \qquad \text{(III.27).}$$

En ce qui concerne la mesure de performance coté serveur, citons la probabilité que le serveur soit occupé, c'est-à-dire la probabilité qu'il y ait au moins un client dans le système, soit :

$$P[N \geq 1] = 1 - P[N = 0] = 1 - (1 - q) = q$$

Pour une file d'attente M/M/1, la quantité $q = \frac{\lambda}{\mu}$ représente donc la proportion du temps où le serveur est occupé, ce qui explique l'appellation " taux d'utilisation" donnée précédemment au facteur q.

Enfin, le débit d de la file d'attente est le nombre de sorties par unité de temps, soit le taux de service (µ) multipliée par la proportion du temps où il y a effectivement une sortie, c'est-à-dire la proportion du temps où le serveur est occupé. On obtient :

$$D = \mu P(N \geq 1) = \mu . q = \lambda ;$$

Autrement dit, le débit de la file est égal au taux d'arrivée! C'est un résultat intuitivement correct pour une file d'attente stable : à l'équilibre, le taux de sortie de la file est égal au taux d'arrivée.

Le tableau suivant résume les formules servant à calculer les mesures de performance pour un modèle avec serveur unique M/M/1 :

Mesure de performance	**Equation**
Nombre moyen de clients en file	$\tilde{n}_f = \frac{q^2}{1-q}$
Nombre moyen de clients dans le système	$\tilde{n} = \frac{q}{1-q}$
Temps moyen d'attente en ligne	$T_f = \frac{1}{\mu}\,\tilde{n} = \frac{q^2}{\lambda(1-q)}$
Temps moyen passé dans le système	$T_s = \frac{1}{\mu-\lambda} = \frac{q}{\lambda(1-q)}$
Probabilité qu'il y ait zéro unité dans le système	$P_0 = 1-q = 1-(\frac{\lambda}{\mu})$
Probabilité qu'il y ait n unités dans le système	$P_n = (\frac{\lambda}{\mu})^n P_0 = (1-q)\,q^n$
Probabilité qu'il y ait moins de k unités dans le système	$P(n \geq k) = q^k$

Tableau III.3 : Formules servant à calculer les mesures de performance pour le système M/M/1

- Variation des mesures de la qualité en fonction du facteur d'utilisation *q*

En plus des valeurs des quantités précédentes, il est intéressant de savoir leur comportement en fonction de la variation du coefficient d'utilisation q. Les graphes III-12 et III-13 montrent que le nombre moyen de clients dans le système ainsi que le temps moyen passé par un client dans le système croient rapidement à mesure que q approche de l'unité. On commettrait donc une erreur si on voulait faire fonctionner un système à sa capacité maximum (en acceptant un taux de service µ minime) : les caractéristiques de la qualité du service se détérioreront gravement.

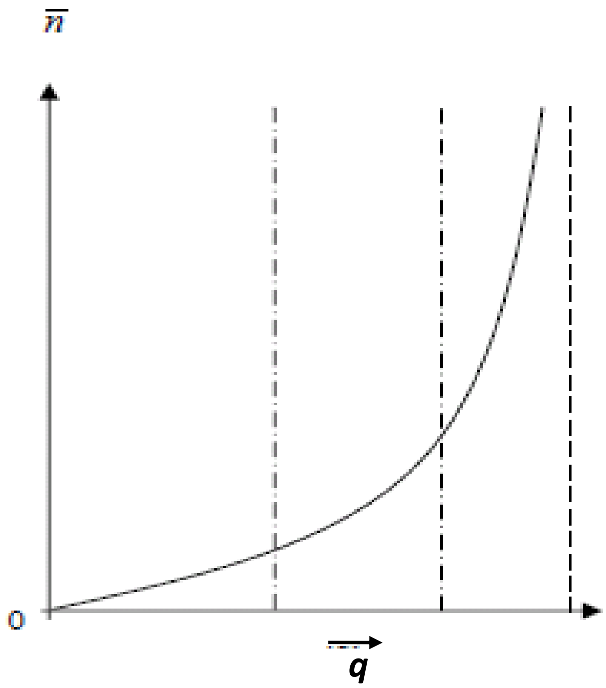


Figure III.10 Variation du nombre moyen n de clients dans un système M/M/1 en fonction du coefficient d'utilisation, q.

On distingue trois phases de l'accroissement des deux courbes de ñ et ū; si la dernière est préjudiciable pour la qualité du service, la première grèverait inutilement le coût de l'exploitation. La phase intermédiaire est donc la phase à rechercher par le gestionnaire.

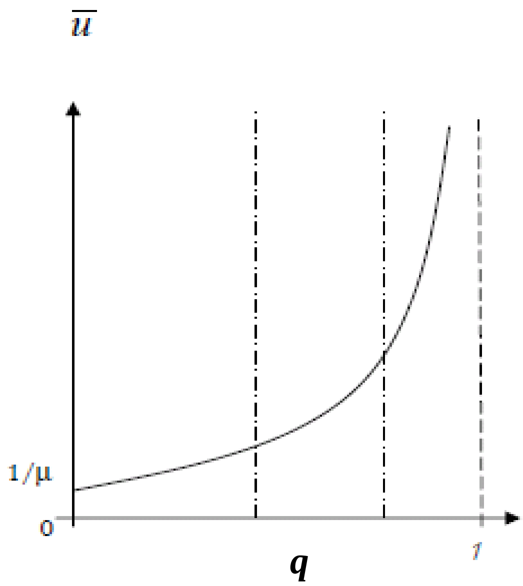


Figure III.11 Variation du temps moyen ū passé par chaque client dans un système M/M/1 en fonction du coefficient d'utilisation, q.

III.8 Conclusion

Dans ce chapitre nous avons présenté les notions de base exprimant les phénomènes d'attente. Généralement, les clients voient dans l'attente une activité sans valeur ajoutée et, s'ils attendent trop longtemps, ils associent cette perte de temps à une mauvaise qualité de service. Les Réseaux de Files d'Attente (RFA) ont une très grande importance en recherche opérationnelle. Ils servent à modéliser et d'analyser des systèmes physiques de type clients/serveurs. Ils permettent ainsi d'évaluer les performances et de mieux comprendre le comportement de ces systèmes.

Nous avons aussi défini dans ce chapitre les éléments importants à considérer dans la mesure de performance d'un système de file d'attente, et on peut dire que le nombre moyen de clients qui attendent en file est l'élément clé qui sert à déterminer les autres mesures de performance du système.

Chapitre IV :

Utilisation de la théorie des files d'attente pour la commande optimale d'un carrefour à feux

IV.1 Introduction

Dans ce chapitre nous exploitons la théorie de files d'attente pour le développement des modèles mathématique d'aide à l'analyse de performance et à l'optimisation des files d'attente dans un carrefour à feu. Ce dernier joue le rôle du routage des flux, venant des différentes lignes. Il constitue un point sensible dans la circulation routière, il est touché d'une manière directe par la congestion du trafic. Il s'avère donc important de réguler des paramètres qui gèrent le trafic dans un carrefour pour fluidifier les flux du réseau routier. Pour évaluer la performance d'un système, on utilise soit les méthodes analytiques, telles que les réseaux de files d'attentes, soit la simulation. Chacune de ces méthodes comporte ses avantages et ses inconvénients.

Nous présentons d'abord les différentes techniques d'évaluation des performances d'un système. En se basant sur les relations qui gèrent le retard pendant la sursaturation dans un carrefour, nous présentons un modèle en temps discret en termes de files d'attente en fonction des cycles. Puis la minimisation d'une fonctionnelle en terme de files d'attente dans les conditions de sursaturation en fin la simulation de l'évolution des files d'attente dans le système.

IV.2 Les différentes techniques d'évaluation des performances

D'après [Caux 1993], évaluer signifie « déterminer une quantité par le calcul sans avoir recours à la mesure directe ». Ceci suppose que l'évaluation soit effectuée à l'aide d'un modèle qui peut être expérimental, mathématique, de simulation... Evaluer implique également le recours à un objectif et un indicateur de performance qui fournit une donnée quantifiée mesurant l'efficacité du système, donc son aptitude à générer une performance [Berrah 1997].

Les différentes techniques d'évaluation des performances d'un système sont schématisées dans la figure ci-dessous (figure IV.1). Elles peuvent être classées en trois grandes catégories : l'obtention de mesures directes sur le système (par la détermination des valeurs directes des critères de performances), les méthodes analytiques et la simulation (Incera, 2001).

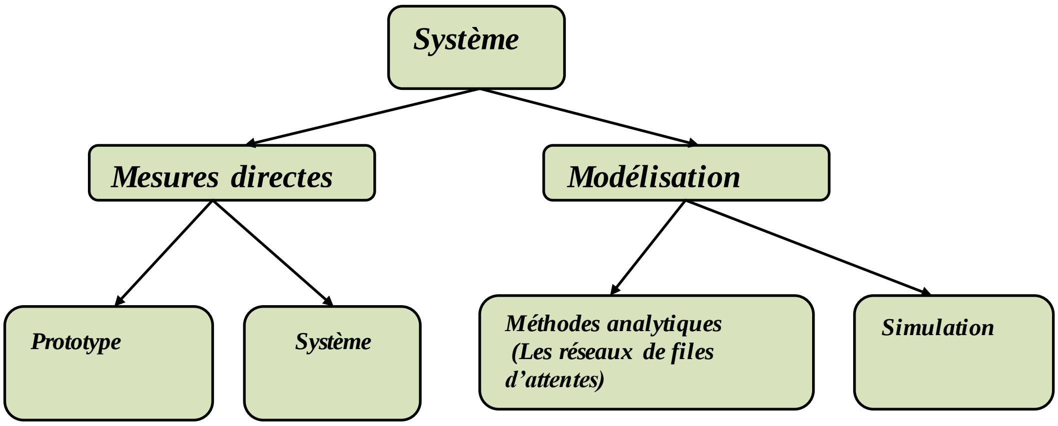


Figure IV.1: Techniques d'évaluation des performances d'un système (Incera, 2001)

La méthode des mesures directes est la seule technique qui peut offrir « l'image réelle » de l'état d'un système réel en tenant compte de toutes les caractéristiques de celui-ci. Elle a, cependant, plusieurs inconvénients. En effet les caractéristiques du système sont extrêmement variables et imprévisibles et par conséquent, les données obtenues à un moment donné sur le système ne permettent pas toujours de prévoir le comportement du système dans d'autres conditions. De ce fait, les méthodes analytiques et la simulation restent les meilleures techniques pour l'évaluation des performances.

Nous allons tout d'abord présenter le système puis le modéliser par un réseau de files d'attente. Cette modélisation nous permet une meilleure compréhension et la formulation mathématique du système. Enfin, nous présentons quelques modèles de simulation comme technique complémentaire pour la méthode analytique RFA.

IV.2.1 Présentation du système

Le système étudié est un carrefour à feux isolé qui comporte deux voies (L1,L2) avec deux feux de signalisation implantés à l'extrémité du carrefour, Figure IV.2. Les voies du carrefour sont caractérisées par un taux d'arrivée moyen égal à Q1 et Q2 et chaque voie du carrefour est de capacité limitée égale à Ca. Le trafic s'écoule dans deux directions: Est -Ouest- (E-O) et Nord-Sud (N-S). Par souci de simplicité, les voies de circulation sont supposées être à sens unique et la prise en compte des mouvements de tournes à gauche et à droite est exclue de notre étude.

Le contrôle des flux de véhicules franchissant un carrefour à feux est mené par des indicateurs de signalisation (vert, jaune, rouge), qui se succèdent à l'intérieur d'un cycle.

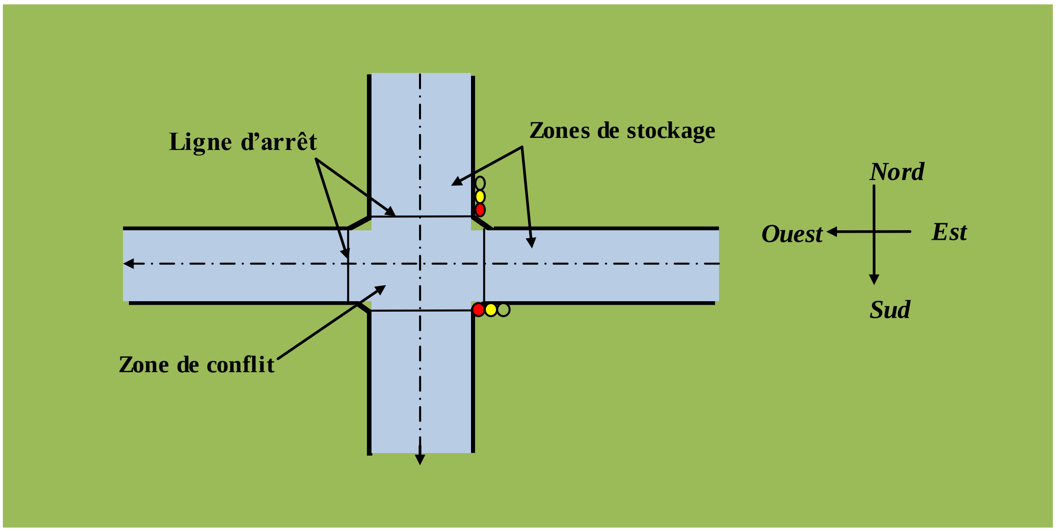


Figure IV.2 : Intersection isolée à deux phases

IV.2.2 Modélisation par files d'attente

L'évaluation des performances d'un système réel se compose d'une étape de modélisation permettant de passer du système au modèle et d'une étape d'analyse des performances du modèle. Nous exposons d'abord les objectifs de la modélisation puis nous modélisons le système par un réseau de files d'attente.

IV.2.2.1 Les objectifs de la modélisation

La modélisation des systèmes réels, née des besoins issus de l'accroissement de leurs complexité , est une discipline qui a pour objectif de pouvoir modéliser les systèmes complexes en tenant compte des nombreux facteurs entrant enligne de compte dans leurs comportements réels. Ainsi le recours à la modélisation est caractérisé par différents objectifs.

Pour Vernadat (1996), le recours à la modélisation des systèmes de biens et de services se justifie par les raisons suivantes :

- comprendre et analyser la structure et le fonctionnement du système ;
- prévoir le comportement et les performances des processus opérationnels et de soutien avant leur implantation ou pendant un projet de réingénierie en vue d'accélérer les flux et d'augmenter la compétitivité du système décrit ;
-choisir la ou les meilleures alternatives de réalisation et d'implantation de toute organisation ;

- bâtir une vision commune du fonctionnement du système et la communiquer facilement au plus grand ensemble possible du personnel.

Ainsi, la modélisation d'entreprise selon F. Verdanat (1999) « a pour objet la construction de modèles d'une partie déterminée d'une entreprise pour en expliquer la structure, le fonctionnement et en analyser le comportement ». Ainsi, l'un des objectifs essentiels de la modélisation d'entreprise est de produire de la connaissance sur une entreprise pour permettre une analyse.

Dans notre cas, le réseau de files d'attentes est constitué de deux files d'attente M/M/1 entre lesquels circulent des flots de voitures. Le réseau est ouvert et comprend un seul serveur (carrefour) qui offre le service. L'arrivée des clients est Poissonienne de taux Q (nombre moyen de clients arrivant pendant une unité de temps) et la durée du service est exponentielle de taux S (nombre moyen de clients servis pendant une unité de temps). Les deux files sont donc considérées comme un processus de Markov en entrée et en sortie, avec un seul serveur (carrefour), une discipline de service premier arrivé, premier servi, une capacité infinie et un nombre infini de clients qui peuvent entrer dans les deux files.

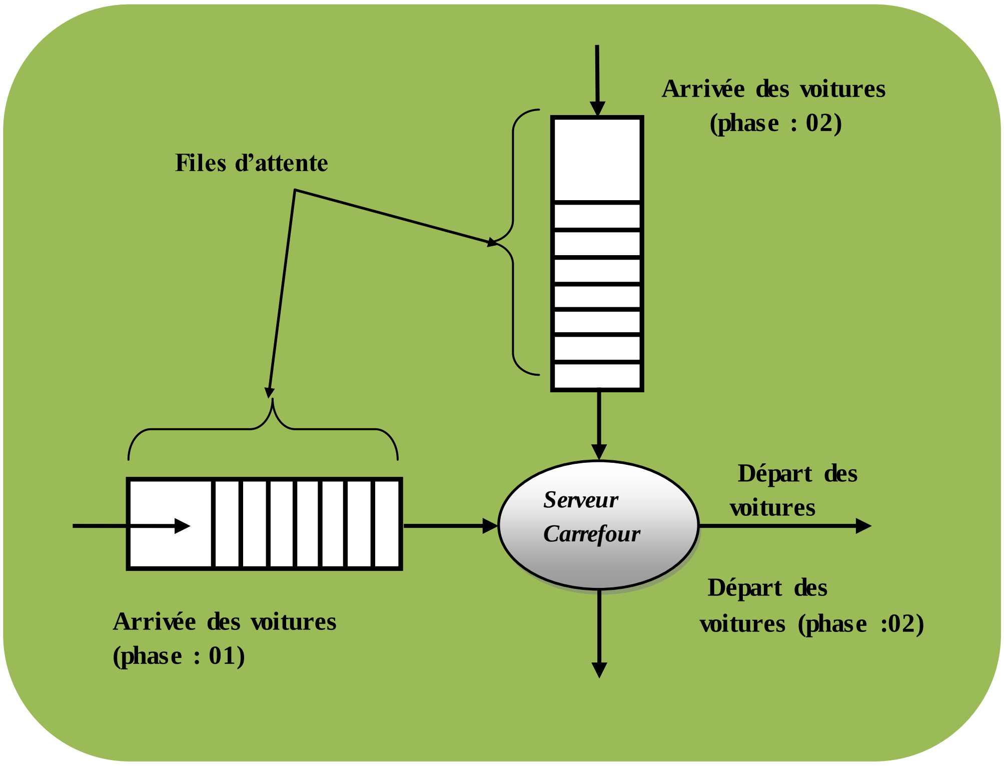


Figure IV.3: modélisation du système par un réseau de files d'attente

IV.2.3 Les équations mathématique régissant le système

Avant la modélisation mathématique du système nous expliquons d'abord certain nombre de notions permettent de décrire les carrefours à feu.

1) Cycle : le cycle peut être défini comme étant la durée constante séparant deux passages successifs de l'ensemble des signaux. Dans un carrefour simple, le cycle est partagé en deux phases qui s'expriment par le temps pendant lequel un ou plusieurs courants sont admis dans le carrefour.

2) Conflit : croisement de deux mouvements, de véhicules ou de piétons.

3) Courant : ensemble de mouvements réunis sur une même voie.

4) Condition de sursaturation: Une phase est dite saturée lorsqu'un véhicule au moins est contraint d'attendre plus d'un cycle pour franchir le carrefour. Le carrefour est dit saturé quand au moins une de ses phases est saturée. Plus formellement, soit C la longueur d'un cycle et notons A(k) comme étant le nombre des arrivées durant un cycle $K \in N$, alors A(k) peut s'exprimer de la manière suivante :

$$A(k) = A[kc] - A[(k-1)c] \quad (IV.1)$$

où A[kc] défini le nombre des arrivées à la fin du cycle k . Soit D_k (c) le nombre des départs pendant le cycle k , alors la condition du sursaturation est définie par :

$$A(k) > D(c) \qquad (IV.2)$$

-Temps effectif du feu vert: noté Ge est le temps réel pendant lequel les véhicules franchissent réellement la phase d'un carrefour (Figure IV.4). Il est défini par l'équation suivante [Webster 1958]:

$$Ge = G + Y - (W1 + W2) \qquad (IV.3)$$

où:

$-W1$: le temps dû au retard d'accélération des véhicules pendant le début du feu vert réel G.

$-W2$: le temps dû au retard de ralentissement des véhicules pendant la fin du feu jaune réel Y.

Figure IV.4 : Temps effectif du vert.

-Taux de saturation: A l'entrée d'un carrefour à feux et pendant le temps effectif du feu vert Ge, le taux de saturation S est défini comme étant le nombre maximum de véhicules pouvant utiliser le couloir sans interruption.

IV.2.3.1 formulation mathématique du système

Dans cette section, nous expliquons, dans une première partie, l'équation récurrente qui représente, d'une manière générale, la relation entre les longueurs des files d'attente durant une succession de cycles pendant la période de sursaturation. Dans une deuxième partie, nous présentons le système par un model discret grâce une nouvelle formulation sur les files d'attente. Le choix de ce modèle discret est dû à la cohérence entre les périodes dans lesquelles évoluent le système et les changements des signalisations. Ces derniers se produisent exactement à l'arrêt d'un cycle.

Le system se rapporte facilement à un processus de naissance et de mort où X(t) représente le nombre de clients dans le système, chaque arrivée est considérée comme une naissance et chaque fin de service est considérée comme une mort.

Soit $L(k)$ la longueur de la file d'attente dans un couloir à la fin du cycle k. L'équation récurrente qui gère les files d'attente pour une succession des cycles est :

$$L(k+1) = L(k) + A(k+1) - D(k+1) \qquad (IV.4)$$

Pendant le temps effectif du feu vert Ge et en tenant compte de la condition de sursaturation, le nombre de départs est supposé constant et est représenté dans chaque cycle par D(c) (Figure IV.5):

$$D(k+1) = D(c) = S.Ge \quad (IV.5)$$

L'équation précédente (IV.4) s'écrit alors sous la forme:

$$L(k+1) = L(k) + A(k+1) - S.Ge \quad (IV.6)$$

IV.2.3.2 Représentation d'état discrète du système

Maintenant, nous nous intéressons au model concernant une intersection routière isolée à deux phases. Nous présentons le système grâce une nouvelle formulation sur les files d'attente. Nous supposons que le taux des arrivées clients Q_i soit constant. Nous avons illustré sur (la Figure IV.6) l'évolution des deux files d'attente en prenant en considération le déphasage existant entre les deux lignes.

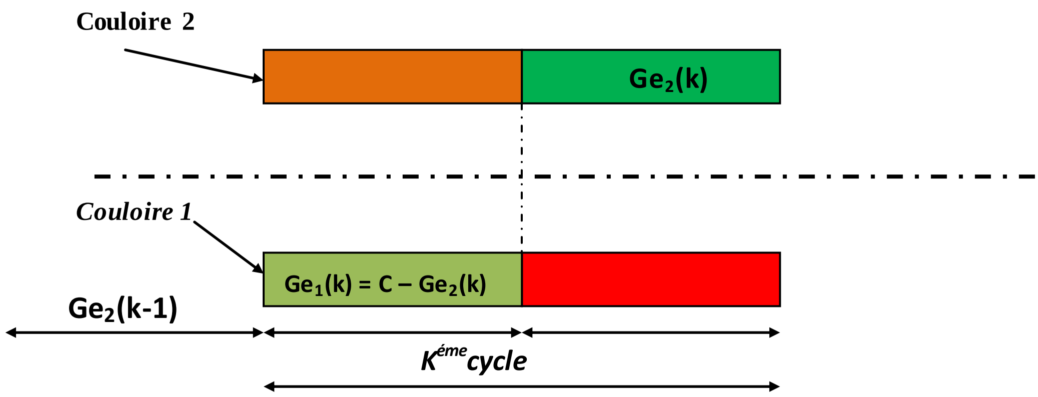


Figure IV.5 : Carrefour à deux phases

Soit $X_2(k)$ la longueur de la file d'attente dans le couloir 2 à la fin du $Ge_2(k)$

L'application de la relation $(IV.4)$ définie dans la section précédente permet d'écrire :

$$X_2(K+1) = X_2(K) + A_2(K+1) - D_2(K+1) \qquad \text{(IV.7)}$$

où,

$A_2(K+1) = Q_2.C$ est le nombre des arrivées à la fin du cycle k.

$D_2(K+1)=S_2.Ge_2(k)$ est le nombre des départs à la fin du cycle k .

Il vient alors:

$$X_2(K+1) = X_2(K) + Q_2.C - S_2.Ge_2(k) \qquad (IV.8)$$

De même, soit $X_1(K)$ la longueur de la file d'attente dans le couloir 1 à la fin du $Ge_1 = C -$ $Ge_2(k)$

On a alors:

$$X_1(k+1) = X_1(k) + A_1(k+1) - D_1(k+1) \qquad \textit{(IV.9)}$$

où,

$A_1(K+1) = Q_1.Ge_2(k-1)+ Q_1.(C- Ge_2(k))$ représente le nombre des arrivées à la fin de $C - Ge_2(k)$.Le terme $Q_1Ge_2(k-1)$ correspondant au nombre des arrivées quand le feu est rouge à la fin du cycle (k-1)de la première phase.

$D_1(K+1) = S_1(C- Ge_2(k))$ représente le nombre des départs à la fin de $(C- Ge_2(k))$

Ainsi, l'équation (IV.9) devient en fonction uniquement de la variable Ge_2:

$$X_1(K+1) = X_1(K)+Q_1Ge_2(k-1) + Q_1(C - Ge_2(k)) - S_1.(C- Ge_2(k)) \qquad \text{(IV.10)}$$

La forme de la file d'attente $X_1(K)$ se justifie de la manière suivante : à la fin de $Ge_1(k)$ le nombre de véhicules présents dans le couloir 1 est égal au nombre de véhicules restants à la fin de $Ge_1(k-1)$, noté $X_1(K)$, auquel on ajoute le nombre des arrivées pendant $Ge_2(k-1)$ et $Ge_1(k)$, moins le nombre des départs pendant $Ge_1(k)$. Nous notons en particulier que cette formulation impose à la ligne 1 d'être en déphasage avec la ligne 2.

Soit maintenant $X(k)$ le vecteur d'état du système défini par $X(k) = (X_1(k)\ \ X_2(k))^T$ et définissons $U(K) = (Ge2(k)\ \ Ge2(k-1))^T$ comme étant le vecteur de commande du système. Les équations (IV.8),(IV.10), peuvent alors s'écrire sous la forme matricielle suivante:

$$X(k+1) = A.X(k) + B.U(k) + C \qquad (IV.11)$$

Avec $X(k) = \begin{bmatrix} X1(K) \\ X2(K) \end{bmatrix}$ $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $B = \begin{bmatrix} S1 - Q1 & Q1 \\ -S2 & 0 \end{bmatrix}$

$$U(k) = \begin{bmatrix} Ge2(k) \\ Ge2(k-1) \end{bmatrix} \qquad C = \begin{bmatrix} (Q1 - S1).C \\ Q1.C \end{bmatrix}$$

IV.2.4 Recherche d'une solution optimale du système

Avant la résolution mathématique du système nous exposons d'abord le principe du minimum de Pontriaguine puis on définit la stratégie de la commande optimale du système.

IV.2.4.1 Le principe du minimum de Pontryagin

Le principe du minimum de Pontryagin est utilisé dans la théorie du contrôle optimal pour trouver la commande optimale permettant d'amener un système dynamique d'un état à un autre, en présence de contraintes portant sur l'état ou les commandes d'entrée.

Le principe examine la minimisation d'un Hamiltonien sur U, l'espace des commandes admissibles. Si $u^* \in U$ est la commande optimale pour le problème, alors le principe énonce que :

$$H(x^*(t),u^*(t),\lambda^*(t),t) \leq H(x^*(t),u(t),\lambda^*(t),t) \qquad \forall u \epsilon U, \quad t \epsilon [t_{0\,,}t_f]$$

où $x^* \epsilon\, C^1\left[t_{0\,,}t_f\right]$ est la trajectoire d'état optimale et $\lambda^* \epsilon\, BV\left[t_{0\,,}t_f\right]$ la trajectoire de co-état optimale.

Ce résultat a été initialement appliqué pour la résolution de problèmes de minimisation de temps de transformation avec contraintes sur les commandes d'entrées, mais il peut également être utilisé pour résoudre des problèmes à contrainte d'état.

Il est également possible de dériver des conditions spécifiques sur l'Hamiltonien. Si l'instant final t_f est fixé et que l'hamiltonien ne dépend pas explicitement du temps ($\frac{\partial H}{\partial x} = 0$),alors :

$$H(x^*(t),u^*(t),\lambda^*(t),t) = \text{constant}$$

si l'instant final n'est pas fixé, alors :

$$H(x^*(t),u^*(t),\lambda^*(t),t) = 0$$

-Les conditions nécessaires de résolution du problème de minimisation

Les conditions nécessaires pour la minimisation d'une fonctionnelle sont les suivantes. Soit x l'état du système dynamique et u la variable de commande, telle que :

$\dot{x} = f(x,u)$, $x(0) = x_0$, $u(t) \epsilon\, U$, $t \epsilon\, [0,T]$ où U est l'espace des commandes admissibles et T la date de l'état final du système. La commande $u(t) \epsilon\, U$ doit être déterminée pour tout $t \,\epsilon\, [0,T]$ afin de maximiser la fonctionnelle J, définie par :

$$J = \psi\, [x(T)] + \int_0^T L[x(t),u(t)]dt$$

Les contraintes sur la dynamique du système peuvent être adjointes au Lagrangien L en introduisant le vecteur des multiplicateurs de Lagrange fonction du temps λ. Ces éléments sont appelés co-états du système.

Cela permet de construire l'Hamiltonien H défini pour tout $t \epsilon\, [0,T]$ par :

$$H\, [\lambda(t),x(t),u(t),t] = \lambda^T(t)\, f[x(t),u(t)] + L[x(t),u(t)]$$

où $\lambda^T(t)$ est le transposé de $\lambda(t)$.

Le principe du minimum de Pontryagin énonce que la trajectoire d'état optimale x^*, la commande optimale u^* , et le vecteur des multiplicateurs de Lagrange correspondant λ^* doivent minimiser l'hamiltonien H de façon à ce que :

$$H(x^*(t),u^*(t),\lambda^*(t),t) \leq H(x^*(t),u(t),\lambda^*(t),t) \qquad \text{pour tout } t \in [0,T]$$

IV.2.4.2 Minimisation des Files d'attente du système

Comme pour toute stratégie de commande optimale, il est nécessaire de définir clairement les points suivants:

- Le critère d'optimisation.
- Le domaine d'optimisation admissible.
- La stratégie de commande proprement dite.

Le critère d'optimisation précise les objectifs par rapport auxquels l'optimalité est définie. Dans notre travail nous nous intéressons à la minimisation des critères suivants :

J_1: la somme des files d'attente pendant une période de sursaturation:

$$J_1 = \sum_{k=0}^{N-1} X(k)$$

J_2 : la forme quadratique des files d'attente:

$$J_2 = \sum_{k=0}^{N-1} 1/2\, x^2(k)$$

Il y va de soi que la définition des critères à minimiser est inspirée par les objectifs de la commande. Or, comme l'un des objectifs de la commande de notre travail est de ramener le système vers un état de non saturation, on comprend aisément que les critères soient définis en fonction des files d'attente.

Le domaine d'optimisation est le sous-ensemble de l'espace des commandes dans le quel une minimisation du critère est à rechercher. Ce sous-ensemble est défini par une limitation d'ordre technique, c'est la nécessité de prendre en compte des contraintes sur la commande et sur l'état. Dans notre cas nous définissons l'ensemble par:

$$W = \{(U(K), X(K)) \in R^2.\, R^2 / U_{min} \leq U(K) \leq U_{max}, X(K) \geq 0\} \quad \text{(IV.12)}$$

Notons que ces deux contraintes traduisent une cohérence avec le fonctionnement réel d'un carrefour. En effet, puisque le cycle de signalisation est borné, le temps effectif du feu vert doit être borné. En outre, la file d'attente traduit dans notre cas le nombre de véhicules présents dans un couloir, nous comprenons aisément qu'elle soit positive.

Une fois le critère et le domaine d'optimisation donnés, la stratégie de commande consiste à choisir connaissant l'état initial $(X_{01}, X_{02})^T$, une suite de commandes admissibles $U(1),...,U(N)$, pour les contraintes (IV.11) et (IV.12), telle que les critères J_i prennent des valeurs minimales.

Le problème ainsi posé peut être résolu par plusieurs méthodes. Pour notre part, nous exploitons la méthode du principe du maximum. En effet, définissons l'Hamiltonien suivant:

$$H(k) = \varphi(k) + \lambda^{T}(k+1).[X(k) + B\,U(k) + C]$$

où φ (k) est la fonctionnelle qui apparaît dans la somme du critère et λ (k) est le multiplicateur de Lagrange. La première tâche est de garantir l'existence de solution au problème:

$$\min_{U_{\min} \le U \le U_{\max}} H(k)$$

L'application du principe du maximum relative à la commande $U(k)$ conduit immédiatement aux équations adjointes suivantes :

$$\frac{\partial H}{\partial X(k)} = \lambda(k), \quad \frac{\partial H}{\partial \lambda(k+1)} = X(k), \quad \frac{\partial H}{\partial U(k)} = 0$$

Puisque la fonction $H(k)$ est une fonction linéaire en la commande $U(k)$, alors l'Hamiltonien conduit naturellement à des commandes de type bang-bang. Ainsi, puisque : $\frac{\partial H}{\partial U(k)} = \lambda^{T}(k+1)B$

La stratégie de commande est définie par :

$$U(k) = Umin \quad si\ \lambda^{T}(k+1)B > 0$$

$$U(k) = Umax \quad si\ \lambda^{T}(k+1)B < 0$$

IV.2.5 La simulation

La simulation de trafic routier est un secteur en pleine expansion. Elle permet de prévoir les flux de véhicules et d'en informer quotidiennement les usagers. Elle peut aussi être une aide à la décision pour les collectivités qui souhaitent construire ou modifier des aménagements.

- **Définition**

La simulation se définit comme l'utilisation ou la résolution de modèles correspondant à un système donné pour étudier le comportement de ce dernier dans un contexte précis. La simulation est couramment utilisée comme outil d'aide à la décision dans le domaine des systèmes de production de biens ou de services [6].

La simulation est l'utilisation ou la résolution de modèles correspondant à un système donné pour étudier le comportement de ce dernier dans un contexte précis. Il s'agit de suivre une démarche [6].

La démarche de simulation selon Drogoul (1993) passe donc par trois étapes distinctes : l'étape de modélisation, qui consiste à construire le modèle du phénomène à étudier, l'étape d'expérimentation, qui consiste à soumettre ce modèle à un certain type de variations, et l'étape de validation, qui consiste à confronter les données expérimentales obtenues avec le modèle à la réalité.

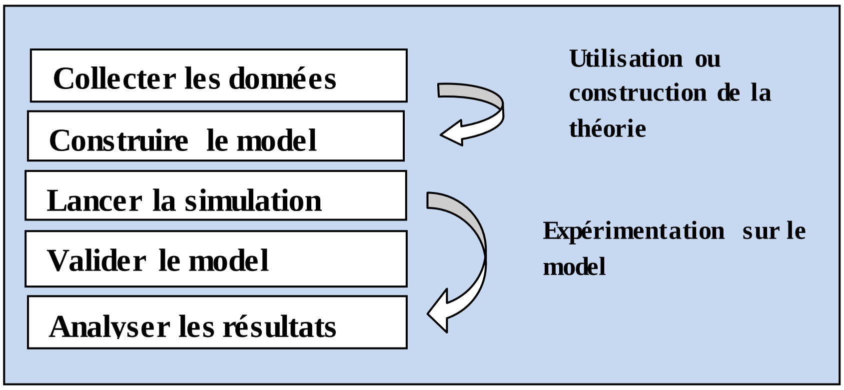


Figure *IV.6*: Les étapes du processus de simulation (Drogoul, 1993)

Dans cette section nous exploitons le logiciel Matlab pour la simulation des lois exprimant le processus des arrivées et de service, ensuite nous présentons le model qui nous aide à prévoir la longueur de la file d'attente dans un carrefour à feu. Notre choix du Matlab est exprimé par les raisons suivants :

- la programmation facile,
- la continuité parmi les valeurs entières, réelles et complexes,
- la gamme étendue des nombres et leurs précisions,
- la bibliothèque mathématique très compréhensive,
- l'outil graphique qui inclus les fonctions d'interface graphique et les utilitaires,
- la possibilité de liaison avec les autres langages classiques de programmations

1) Processus des arrivées

Les arrivées des voitures sont caractérisées par l'ensemble des instants d'arrivée de chaque voiture. Cette collection des dates d'arrivée s'appelle le processus des arrivées. Pendant que les instants d'arrivées sont imprévisibles, elles sont modélisées par des variables aléatoires, et le processus des arrivées est alors une collection de variables aléatoires, c'est-a-dire un processus stochastique (processus de Poisson).

$$P_k(t) = \frac{(\lambda_t)^k}{k!} e^{-\lambda t}$$

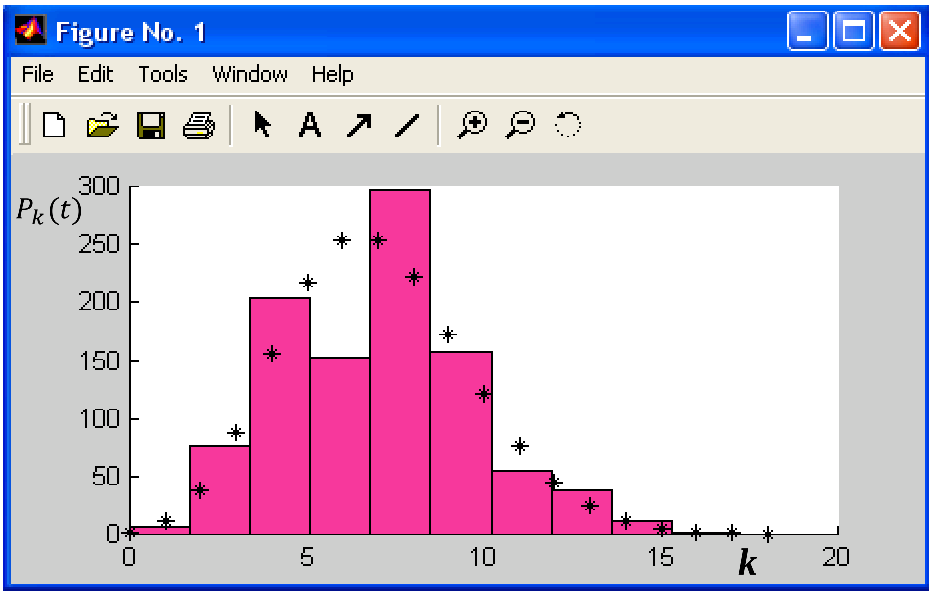


Figure IV .7 : Simulation du processus des arrivées par MATLAB

2) Processus de service

Le processus de service est modélisé par la loi exponentielle de paramétre λ

$$f_x(x) = \lambda e^{-\lambda x}$$

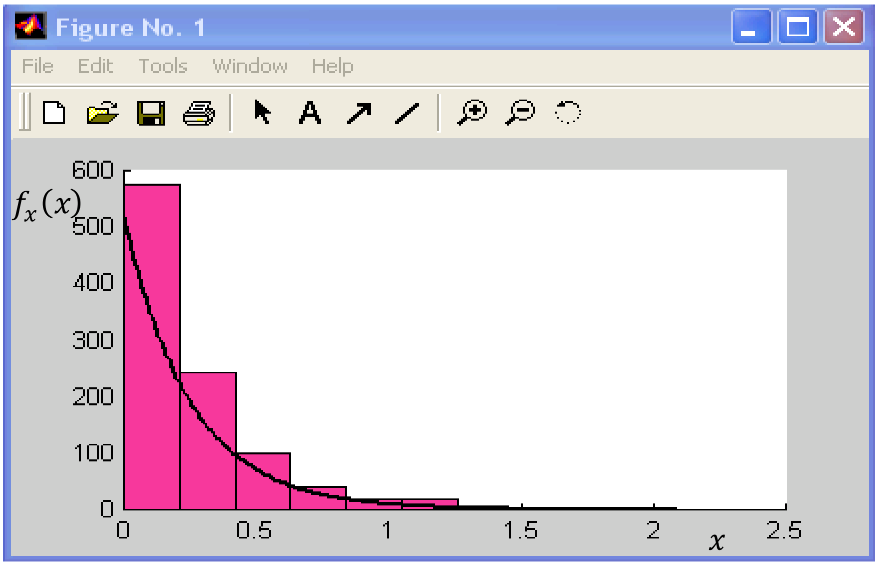


Figure IV .8 : Simulation du processus de service par MATLAB

Les deux modèles de simulation sont probabilistes ; ils permettent d'étudier le comportement temporel du système dont certains paramètres structurels sont donnés sous forme de loi de probabilité.

2) La longueur de file d'attente dans le système

Maintenant, nous exploitons le Matlab pour la simulation des flux des voitures dans le système .Notre objectif est de prévoir la longueur de la file d'attente dans le système dans le dernier cycle, pour cela nous supposons que :

1) Les voitures se déplacent dans une seule direction,
2) la probabilité d'arrivé d'une voiture dans un moment donné est indépendante de l'état précédente du système. Pour l'expérimentation nous supposons que $p = 0.3$,
3) Le temps effectif du feu vert est 20 s
4) Le temps effectif du feu rouge est 20 s
5) Le temps de cycle est 70 s
6) Le nombre des cycles est K=10 cycles

Résultats obtenus

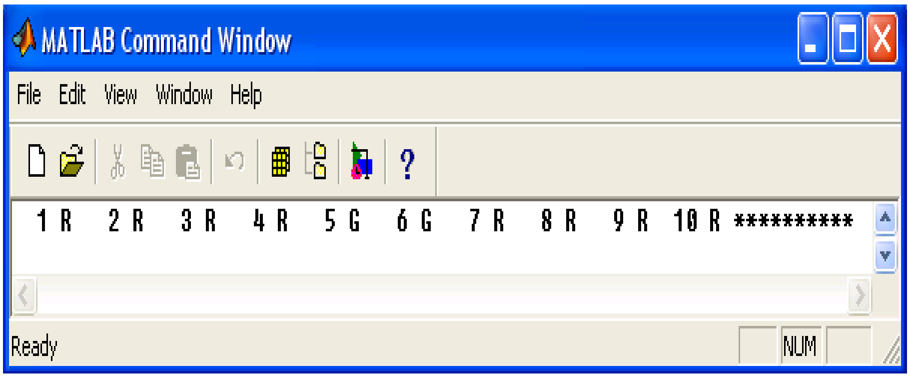


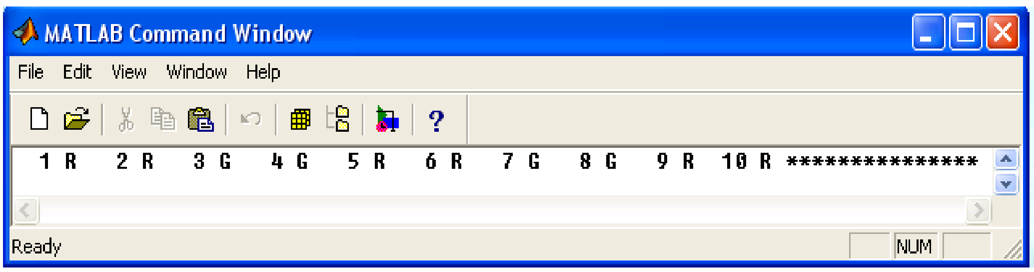
Figure IV.9 : Simulation de la longueur de file d'attente dans le système

Ce modèle de simulation permet de prévoir la longueur de la file d'attente devant un carrefour à feux. Le nombre des étoiles dans le modèle exprime le nombre des clients en attente pour parcourir le carrefour. Les lettres R et G expriment respectivement les couleurs des feux rouge et vert correspondants à chaque cycle K.

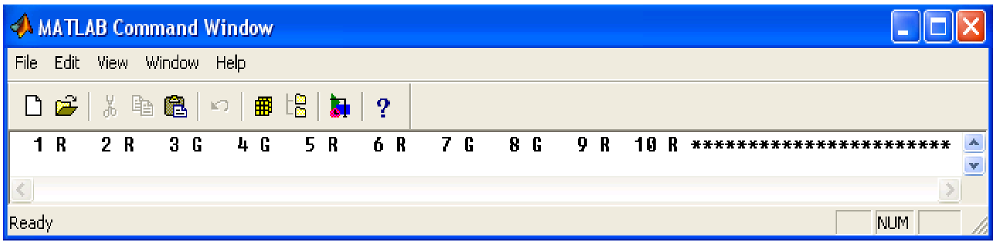
IV.2.5.1 Utilisation du modèle de simulation pour la détermination des variables d'action

Le travail consiste à expérimenter le modèle pour différentes valeurs différentes des temps effectifs des feux rouge et vert et la probabilité d'arrivé d'une voiture dont l'objectif est de déterminer les indicateurs de performance qui influe sur la performance du système.

Premièrement nous fixons le temps effectif du feu vert à 20 s , le temps effectif du feu rouge 20 s et nous augmentons la valeur de la probabilité d'arrivé d'une voiture dans un moment donné , par exemple $p = 0.5$. Le résultat obtenu est comme le suivant :

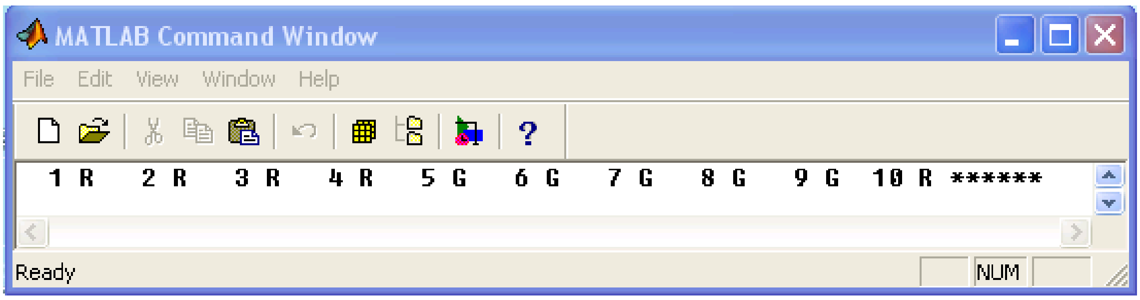


Nous augmentons toujours la valeur de la probabilité d'arrivé d'une voiture dans un moment donné, nous supposons maintenant que $p = 0.6$ et le résultat sera comme le suivant :



Les résultats obtenus montrent qu'avec les conditions précédentes la longueur de la file d'attente est proportionnelle avec a probabilité d'arrivé d'une voiture.

Maintenant nous fixons la valeur de la probabilité d'arrivé d'une voiture à $p = 0.3$, et nous augmentons la valeur du temps effectif du feu vert à 50 s ; et la valeur du temps effectif du feu rouge à 40 s. Le résultat sera comme le suivant :



Nous constatons qu'avec les conditions précédentes la langueur de la file d'attente se diminue. Il est suffisant donc de changer les valeurs de temps effectif des feux vert et rouge pour connaitre la longueur de la file d'attente dans n'import quelle cycle. Ces deux variables sont donc des variables d'action sur lesquelles on peut agir, et peuvent être considérés comme des indicateurs de performances qui permettront aux responsables du service de se rapprocher de l'objectif fixé tel que la minimisation de la longueur de file d'attente dans le système et donc l'amélioration de la qualité de service. De ce fait, il est important de fixer les inducteurs de performance sur lesquels nous pouvons agir et aussi de choisir les axes d'amélioration.

IV.3 Conclusion

Le recours à l'évaluation de performance d'un tel système d'attente est nécessaire avant toute prise de décision, l'objectif étant l'optimisation de sa performance. Cette évaluation représente la mesure de l'impact d'une décision sur le système ou l'influence de perturbations modifiant son état. Les solutions analytiques bénéficient de temps de résolution très rapides. Les résultats peuvent être immédiats, car ils sont déterminés à partir d'équations mathématiques issues du formalisme emprunté. Toutefois, les modèles doivent être suffisamment simples pour demeurer solubles par voie analytique. Ainsi, la zone de fonctionnement de ces solutions, bien qu'étendue par les méthodes de résolutions approximatives, demeure assez restreinte.

La simulation de trafic routier permet de prévoir les flux de véhicules et d'en informer quotidiennement les usagers. Elle peut aussi être une aide à la décision pour les collectivités qui souhaitent construire ou modifier des aménagements.

Conclusion générale

Pour évaluer la performance d'un système, on utilise soit les méthodes analytiques, telles que les réseaux de files d'attentes, soit la simulation. Chacune de ces méthodes comporte ses avantages et ses inconvénients. Les solutions analytiques bénéficient de temps de résolution très rapides. Les résultats peuvent être immédiats, car ils sont déterminés à partir d'équations mathématiques issues du formalisme emprunté. Toutefois, les modèles doivent être suffisamment simples pour demeurer solubles par voie analytique. Ainsi, la zone de fonctionnement de ces solutions, bien qu'étendue par les méthodes de résolutions approximatives, demeure assez restreinte.

L'utilisation des files d'attente peut être un aspect important de la conception des systèmes. Les files d'attente ont tendance à se former, bien que, d'un point de vue macro, les systèmes ne soient pas congestionnés. Les arrivées aléatoires des clients combinées à la variabilité des temps de service créent temporairement des congestions dans le système, d'où la création de files d'attente.

Afin de répondre au mieux au besoin d'amélioration dans un établissement quelconque, il est important de pouvoir proposer une modélisation. Celle-ci va permettre de mieux identifier les dysfonctionnements et problèmes rencontrés. La modélisation constitue une aide considérable au niveau de la phase d'analyse de l'existant.

Le secteur des services se distingue des autres secteurs économiques en ce sens que le personnel en est le principal facteur de production. Il est donc essentiel que le personnel développe et renouvelle ses connaissances afin d'optimiser le potentiel d'emploi qu'offrent les services. Et puisque le personnel constitue, logiquement, le principal intrant, la recherche de qualité doit être centrée sur le personnel.

Dans le domaine dans l'administration publique, offrant essentiellement des services, la démarche qualité vise la plupart du temps à réduire l'incoercible écart entre les attentes ou les besoins des bénéficiaires, l'engagement du prestataire ou son offre, et les prestations effectivement fournies. La diminution de cet écart dépend certes de la capacité du prestataire à mettre en œuvre ce à quoi il s'est engagé, mais aussi de sa capacité à formaliser les attentes – parfois implicites – du bénéficiaire, et à négocier le contenu possible des prestations avec ce dernier. Ces points s'avèrent encore plus essentiels quand il s'agit de services relationnels.

Bibliographies

[1]. AFÊCHE Philipp., -Delay performance in stochastic processing networks with priority service, Operation Research Letters, Sep. 2003, Vol. 31, P. 390-400.

[2]. BAYNAT B., -Théorie des files d'attente : Des chaines de Markov aux réseaux a forme produit, coll. Réseaux et télécommunications, 2000.

BELAID R., –Analyse du phénomène d'attente dans un système d'accostage- essai sur un secteur du port d'Alger, Mémoire de Magister, Directeur de recherche Alexandre LAPCO, INPS, Alger, Juin 1986.

[3]. BOUAZIZ R., -Application de la technique des files d'attente aux problèmes de gestion de l'aéroport de Houari Boumediene, Mémoire de Magister, Directeur de recherche Paul DE MOUNTER, DR, SC. INPS, Alger, 1986.

[4].CARTON D., -Processus aléatoires utilises en recherche opérationnelle, Masson, Paris, 1975.

[5]. DESBAZEILLE G., -Exercices et problèmes de Recherche Opérationnelle, 2eme édition, Dunod, Paris, 1976.

[6]. GELEMBE., -Introduction aux réseaux de files d'attente, 1985.

[7]. KAUFMANN A., -Méthodes et modèles de la Recherche Opérationnelle (les mathématiques de l'entreprise), Tome 1, 2eme édition, Dunod, Paris, 1970.

[8]. PELLAUMAIL. J., -Probabilités statistiques, files d'attente : cours et exercices résolus, 1986.

[9]. QUITTARD P., -Processus stochastiques et file d'attente, OPU, Algérie, 1983.

[10]. ROBERT P., -Réseaux de files d'attente méthodes probabilistes, SPRINGER VERLAG,

Collection MATHEMATIQUES ET APPLICATIONS (SMAI), France, 2000.

[11]. TAKACS L., -Processus stochastiques, problèmes et Solutions, Dunod, Paris, 1964.

[12].Tungsheng Yu (1997), On-line Traffic Signalization using Robust Feedback Control.Thèse, Faculty of the Virginia Polytchenic Institue.

[13]. B.De Schutter and B.De Moor (1998), Optimal traffic light control for a single intersection. European Journal of Control, vol. 4, no. 3, pp.260-276.

[14].R.K.Boel and Bart De Schutter(1999), Approaches to modelling, analysis, and control of hybrid system, European Journal of Control, vol. 40, no.4.

[15]. Akçelik (1980), Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic. Vermont South, Victoria: Australian Road Research Board.
[16]. (1958), Traffic Signal Settings. London: Her Majesty's Stationery Office.
[17].Tapio Luttinen, Riku Nevala (2002), Capacity and Level of Service of Finnish Signalized Intersections. Finnish Road Administration, Helsinki 2002.
[18].Claude Yves Bernard, AFNOR(2000), Le management par la qualité totale.
[19].Mohammed NOUIGA, « La conduite du changement par la qualité dans un contexte socioculturel essai de modélisation systémique et application à l'entreprise marocaine » thèse de doctorat 10 janvier 2003.
[20].A. Bartoli. Le management dans les organisations publiques, Ed. Dunod, 1997.

Liste de figures

Liste de tableaux