Files d'attente

Plan du cours :

- Chapitre 1 : Introduction
- Chapitre 2 : Rappels mathématiques
- Chapitre 3 : Les chaînes de Markov
- Chapitre 4 : Files d'attente
- Chapitre 5 : Réseau de files d'attente

Marion Gilson-Bagrel

1

Plan

Plan du chapitre 3 :

- Introduction
- Présentation
- Etude d'une file d'attente
- Exercices

2

Introduction

- **Historique**
  - File d'attente = fondée sur la théorie des processus markovien
    - ➔ 1960 : file d'attente simple
    - ➔ 1960 → 1980 : réseaux de files d'attente à forme produit (solution exacte)
    - ➔ 1980 → … : méthode approximative
- **Application**
  - Système informatique
  - Réseaux de télécommunication
  - Systèmes de production
- **Phénomène d'attente**
  - Des clients *arrivent* à un certain endroit et réclament un certain *service*
  - Instants d'arrivée et durées de service = quantités aléatoires
  - Si poste de service libre, client servi
  - Sinon client dans file d'attente
- **Objectif :** étudier les performances d'un système

3

Plan

Plan du chapitre 3 :

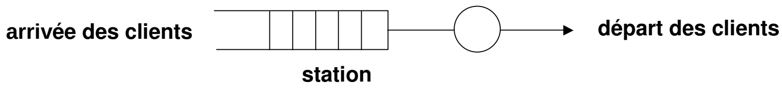
- **Introduction**
- **Présentation**
  - **Définition**
  - **Exemple 1**
  - **Exemple 2**
  - **Outils**
- **Etude d'une file d'attente**
- **Exercices**

4

Présentation

Définition :

- **File d'attente unique :**



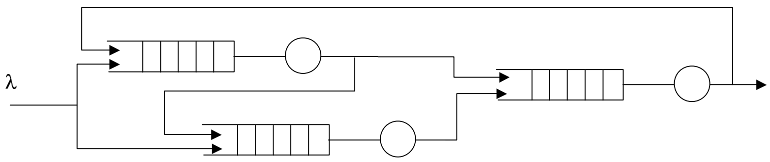
- **Une station :**
  - Composée d'une file d'attente et d'un ou plusieurs serveurs
  - Définie par :
    - ➔ Processus d'arrivée des clients à la station
    - ➔ Discipline de service (PAPS, …) First In, First Out
    - ➔ Nombre de serveur (éventuellement infini)
    - ➔ Processus de service
    - ➔ Capacité de la file d'attente (éventuellement infinie)

5

Présentation

Définition : réseau de file d'attente

- **Ensemble de stations interconnectées**
- **Réseau ouvert (monoclasse) :**



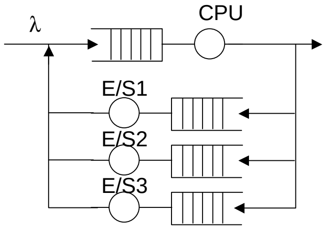
- Nombre de clients dans le réseau variable, en général non borné
- Définition d'un réseau ouvert :
  - ➔ La description de chacune des stations
  - ➔ Le processus d'arrivée des clients dans le réseau
  - ➔ Le cheminement des clients dans le réseau
- Possibilité de définir des réseaux ouverts à capacité totale limitée

6

Présentation

Définition : réseau de file d'attente

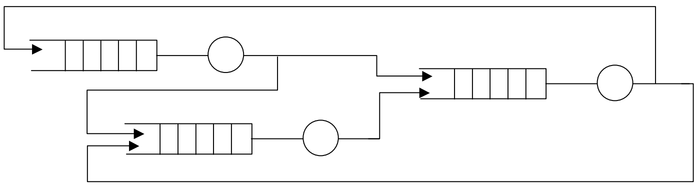
- **Exemple de réseau ouvert (monoclasse) :** modèle serveur central
  - Système =
    - ➔ unité centrale : exécution des processus
    - ➔ ensemble d'unités d'entrée-sortie de stockage : CD, disquette, clé USB, bande magnétique …
    - ➔ Entités circulant dans le système : processus
      - ✓ soit prêt : en attente de libération de l'unité centrale
      - ✓ soit élu : en exécution sur l'unité centrale
      - ✓ soit en attente : attente ou exécution d'une E/S sur une des unités
  - Modélisation :
    - ➔ Réseau ouvert
    - ➔ Clients = processus
    - ➔ Ressources : stations
    - ➔ Entrée : fréquence moyenne λ
    - ➔ Sortie : out ou stockage



Présentation

Définition : réseau de file d'attente

- **Réseau fermé (monoclasse) :**



  - Nombre de clients dans le réseau constant
  - Définition d'un réseau fermé :
    - ➔ La description de chacune des stations
    - ➔ Le cheminement des clients dans le réseau
    - ➔ Le nombre de clients dans le réseau
- **Réseau monoclasse :** même traitement pour chaque client
- **Réseau multiclasse :** traitement différent

Présentation

Définition : réseau de file d'attente

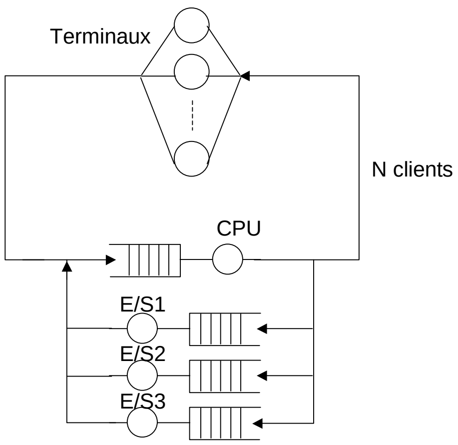
- **Exemple de réseau fermé :** calculateur central avec terminaux
  - Système formé :
    - ➔ d'un calculateur central
    - ➔ d'un ensemble d'unités d'entrée/sortie
    - ➔ d'un ensemble de terminaux
  - Fonctions :
    - ➔ les terminaux génèrent des tâches qui parviennent au calculateur
    - ➔ après l'exécution de la tâche : retour au terminal
    - ➔ l'utilisateur relance périodiquement (après réflexion) une nouvelle tâche
    - ➔ pour l'exécution des tâches : réalisation de certaines E/S
  - Modèle :
    - ➔ clients : tâches
    - ➔ réseau fermé : tâches en nombre fixe, déterminé par le nombre d'utilisateurs (terminaux)
    - ➔ temps de service : temps de réflexion de l'utilisateur

9

Présentation

Définition : réseau de file d'attente

- **Exemple de réseau ouvert fermé :** calculateur central avec terminaux
  - modélisation :



10

Présentation

Définition : classe - chaîne

- **Classe de clients :**
  - Ensemble de clients dont le comportement a la même description :
    - ➔ Temps de service aux stations identiques
    - ➔ Cheminement identique (plusieurs chemins possibles)
  - Type de classe de clients :
    - ➔ Réseau monoclasse (ouvert ou fermé)
    - ➔ Réseau multiclasse : il faut décrire
      - ✓ chaque station
      - ✓ Le cheminement de chaque classe de clients
      - ✓ réseau ouvert : processus d'arrivée de chaque classe
      - ✓ réseau fermé : nombre de client dans chaque classe
  - Possibilité de changement de classe pour un client
- **Chaîne :**
  - Ensemble de classes auxquelles peut appartenir un même client

11

Présentation

Exemple : supermarché

- **Un supermarché est composé de**
  - **Rayons dans lesquels les clients se servent**
  - **1 rayon fromagerie (1 serveur)**
  - **1 rayon charcuterie (1 serveur)**
  - **N caisses**
- **Cheminement des clients :**
  - Les clients du magasin vont :
    - ➔ dans les rayons libre-service
    - ➔ puis au rayon fromage et/ou au rayon charcuterie
    - ➔ à la fin des courses : ils vont payer à l'une des N caisses
  - 2 temps d'attente :
    - ➔ temps d'attente pour être servi
    - ➔ temps où l'on se sert (client devient serveur)
      - ✓ rayon libre-service = file d'attente

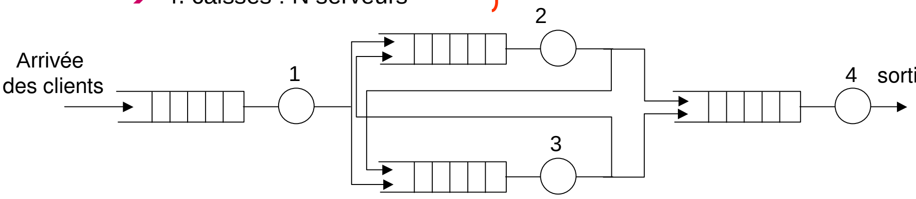
12

Présentation

Exemple : supermarché (suite)

- **Modélisation**
  - 4 stations :
    - → 1. libre-service : serveur infini (SI)
    - → 2. charcuterie : mono-serveur
    - → 3. fromagerie : mono-serveur
    - → 4. caisses : N serveurs

Caractérisation de la file : PAPS



Description du réseau :

- → **Clients du réseau : clients du magasin**
- → **Réseau ouvert, monoclasse : définir**
  - ✓ **processus d'arrivée**
  - ✓ **processus de service**
  - ✓ **cheminement des clients**

Présentation

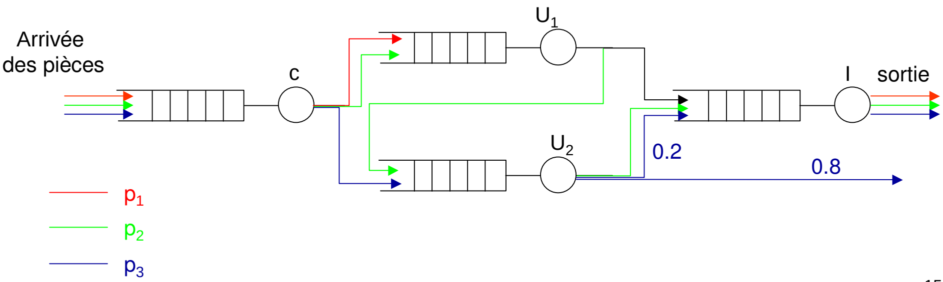
Exemple : système de production

- **Production de 3 types de pièces : p_1, p_2, p_3. Les pièces brutes sont chargées sur des palettes.**
- **Un poste de travail comporte :**
  - **1 chambre de chargement : c**
  - **2 centres d'usinage : U_1, U_2**
  - **1 centre d'inspection : I**
- **Gamme des pièces :**
  - **p_1 : chargement → U_1 → Inspection**
  - **p_2 : chargement → U_1 → U_2 → Inspection**
  - **p_3 : chargement → U_2 → Inspection (pour 1/5)**
- **1er cas à traiter : palettes en nombre suffisant, d'une seule sorte universelle. Question : modélisation du problème.**

Présentation

Exemple : système de production (suite)

- **Modélisation :**
  - Définition du réseau :
    - ➔ Clients du réseau = pièces
    - ➔ Réseau ouvert (car les palettes sont en nombre suffisant)
    - ➔ 3 classes (suivant les 3 pièces différentes)
    - ➔ 3 chaînes (car une pièce ne peut pas changer de classe)
  - Modélisation :

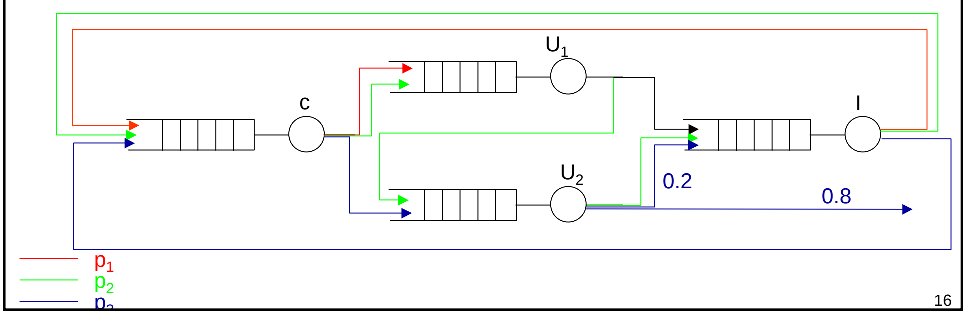


15

Présentation

Exemple : système de production (suite)

- **2ème cas à traiter : palette en nombre déterminé (N), 1 seul type de palette (universel). Question : modéliser le système.**
  - Définition du réseau :
    - ➔ Clients du réseau = palettes (car peut-être en nombre insuffisant)
    - ➔ Réseau fermé
    - ➔ 3 classes (car tout dépend de la pièce placée sur la palette)
    - ➔ 1 chaîne (N clients, d'un seul type)



16

Présentation

Exemple : système de production (suite)

- **3ème cas à traiter :**
  - **2 types de palette (A et B)**
  - **Palettes en nombre déterminé (N_A et N_B),**
  - **Toujours des pièces brutes à l'entrée**
  - **p_1 et p_2 portés par A, p_3 portées par B**
  - **Question : modéliser le système.**

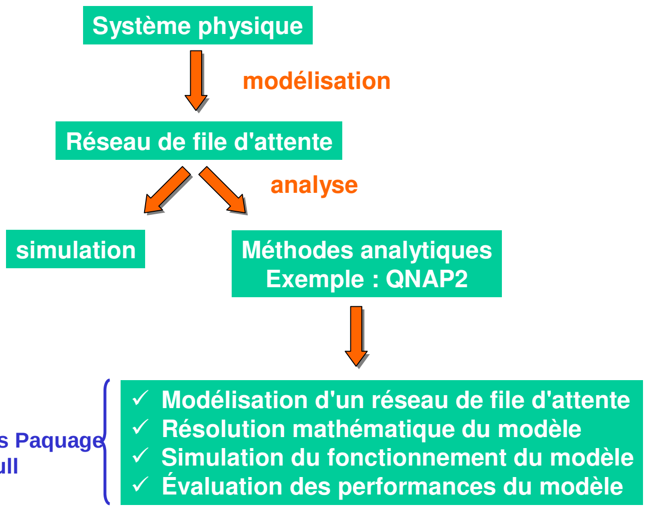
  - Définition du réseau :
    - ➔ Clients du réseau = palettes (car peut-être en nombre insuffisant)
    - ➔ Réseau fermé
    - ➔ 3 classes (car tout dépend de la pièce placée sur la palette)
    - ➔ 2 chaînes (car 2 types de clients : N_A et N_B)

  - Modélisation : identique au cas 2

17

Présentation

Outils :



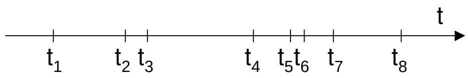
QNAP :
Queing Network Analysis Paquage
dévelopé par INRIA et Bull

18

Présentation

Outils (suite)

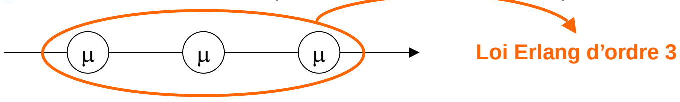
- **Processus de renouvellement :**



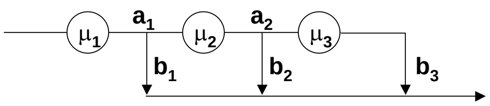
  - Définition : le processus d'arrivée est un processus de renouvellement si les variables aléatoires $T_i=t_{i+1}-t_i$ sont iid (indépendantes et identiquement distribuées)

- **Processus de renouvellement de Poisson :**
  - Processus de renouvellement dont les temps inter-arrivés ($t_{i+1}-t_i$) sont distribués selon un loi exponentielle
  - Probabilité d'une arrivée entre t et dt = λdt

- **Loi Erlang =** Succession de lois exponentielles de taux identiques



- **Loi de Cox =** chaque étage est une loi exponentielle (facile à approcher par toute loi)



$a_i + b_i = 1$

Plan

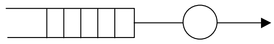
Plan du chapitre 3 :

- **Introduction**
- **Présentation**
- **Etude d'une file d'attente**
  - **Présentation**
  - **Approche stochastique**
  - **M/M/1 + exemple + exercice**
  - **M/M/1/N**
  - **M/M/C**
- **Exercices**

Etude d'une file d'attente

Présentation

- **Principales performances :**
  - X : débit moyen de la file d'attente
  - W : temps de réponse moyen (attente + service)
  - Q : nombre moyen de clients



- **Loi de little :** Q = W X
- **Notation de Kendall :**

A / B / C / K / M / Z

- A : **processus d'arrivée** λ
- B : **processus de service** μ
- C : **nombre de serveurs**
- K : **capacité de la file** (+∞ ss-entendu)
- M : **population des usagers** (+∞ ss-entendu)
- Z : **discipline de service** PAPS (ss-entendu) DAPS …

Pour spécifier les lois de distribution de A et B :

M : exponentiel (Markov) **GI :** distribution générale

E_k **:** Erlang d'ordre k C_k **:** Cox d'ordre k

21

Etude d'une file d'attente

Présentation

- **Exemple :**
  - M/M/2/6 :
    - ➔ A = M : inter-arrivées exponentielles
    - ➔ B = M : service exponentiel
    - ➔ C = 2 : 2 serveurs
    - ➔ K = 6 : capacité de la file, 6 clients
    - ➔ Z non renseigné donc discipline de service par défaut = PAPS

22

Etude d’une file d’attente

Approche stochastique :

- **État :** n(t) clients dans une file donnée à l'instant t
- **Objectif :**
  - Calcul de la probabilité d'avoir n clients : p(n), avec n = 0,1,2,…
  - Déduction des performances :
    - ➔ Nombre moyen de clients $Q = \sum n\,p(n)$
    - ➔ Débit moyen

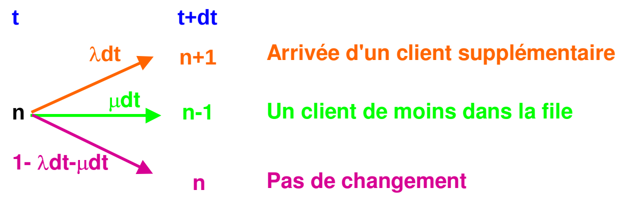
$$X = \underbrace{(1 - p(0))}_{\text{probabilité d'avoir au moins un client}}\ \underbrace{x}_{\text{taux de service}}$$

- **Démarche :**
  - Décrire l'évolution de la file par une chaîne de Markov
  - Résoudre la chaîne de Markov
    - ➔ En déduire p(n) pour n = 1,2, …

Etude d’une file d’attente

Etude de la M/M/1 :

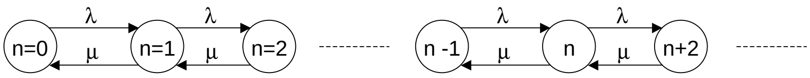
- **définition :**
  - Processus d'arrivée : processus de Poisson de taux λ
  - Processus de service : loi exponentielle de taux μ
  - 1 serveur
  - Capacité infinie
- **n(t) = processus aléatoire markovien**



Etude d'une file d'attente

Etude de la M/M/1 :

- **Chaîne de Markov à temps continu associée:**



  - p(n,t) = probabilité qu'il y ait n clients à l'instant t

    ➔ Pour $n \geq 1$:

$$\underbrace{p(n, t+dt) = p(n, t)(1 - \lambda dt - \mu dt)}_{\text{proba d'aucune évolution}} + \underbrace{p(n-1, t)\,\lambda dt}_{\text{proba de gagner un client}} + \underbrace{p(n+1, t)\mu dt}_{\text{proba de perdre un client (servi)}}$$

- **Régime stationnaire :**
  - ∃ régime stationnaire si p(n,t)=p(n) (indépendant du temps)
  - Définition :

$$\begin{cases} \sum p(n) = 1 \\ p(n) = \rho^n (1-\rho) \end{cases} \quad \text{avec} \quad \rho = \frac{\lambda}{\mu} < 1 \quad \text{et} \quad p(0) = 1 - \rho$$

25

Etude d'une file d'attente

Etude de la M/M/1 :

- **Performances :**
  - Débit :

    ➔ $X = \mu(1 - p(o)) = \lambda$
  - Longueur moyenne de la file d'attente (= nombre moyen de clients)

    ➔ $Q = \sum n\,p(n) = \dfrac{\rho}{1-\rho}$
  - Temps de réponse moyen (ou temps de séjour moyen)

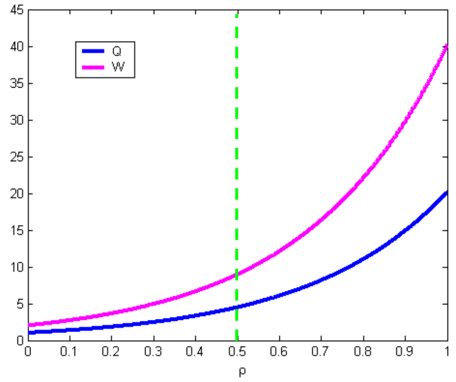
    ➔ $W = \dfrac{Q}{X} = \dfrac{1}{\mu}\dfrac{1}{1-\rho}$
- **Remarque :**
  - λ peut être vu comme : $\lambda = \dfrac{\text{nombre de clients}}{\tau_{\text{arrivée}}}$
  - μ peut être vu comme : $\mu = \dfrac{\text{nombre de serveur}}{\tau_{\text{service}}}$

26

Etude d'une file d'attente

Etude de la M/M/1 :

- **Remarques**



ρ<0.5 : ~ linéaire
(arrivée client= départ client)

ρ>0.5 : tps d'attente augmente
(arrivée client> départ client)

ρ~1 : tps d'attente →∞

Q et W : non linéaires par rapport à ρ

27

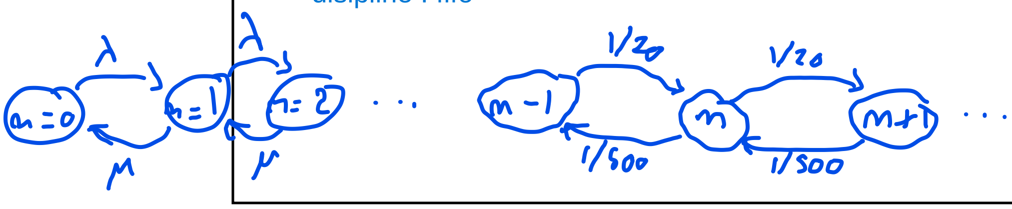
Etude d'une file d'attente

Etude de la M/M/1 : exemple (démonstration des précédente formules)

- **Temps de réponse d'un ordinateur monoprocesseur tel que :**
  - **Temps de service : loi exponentielle, τ_1=500 ms** → $\mu = \frac{1}{500}$
  - **Processus d'arrivée : chaque terminal envoi une requête avec une loi exponentielle de moyenne τ_2= 20s** → $\lambda = \frac{1}{20}$
  - **n terminaux connectés**
- **Question 1 : caractérisation de la file**

procc arrivée : expo de taux 1/20
proc service : expo de taux 1/500
nb serv : 1
capa file : infini
disipline : fifo

→ M/M/1



28

Etude d'une file d'attente

Etude de la M/M/1 : exemple (suite)

- **Question 2 : Matrice de transition**

$$
\begin{array}{c|ccccccc}
 & 0 & 1 & 2 & \cdots & m-1 & m & m+1 \\
\hline
0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\
1 & \mu & -1-\mu & 1 & \cdots & 0 & 0 & 0 \\
2 & 0 & \mu & -1-\mu & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
m-1 & 0 & 0 & 0 & \cdots & -1-\mu & 1 & 0 \\
m & 0 & 0 & 0 & \cdots & \mu & -1-\mu & 1 \\
m+1 & 0 & 0 & 0 & \cdots & 0 & \mu & -1-\mu
\end{array}
$$

29

Etude d'une file d'attente

Etude de la M/M/1 : exemple (suite)

- **Question 3 : équations d'équilibre (ou de balance)**
  - À partir de la matrice de transition, on obtient le système suivant :

-p(0) + up(1) = 0
p(0) - (1+u)p(1) + up(2) = 0
p(1) - (1+u)p(2) + up(3) = 0
...
p(n-2) - (1+u)p(n-1) + up(n) = 0
...

  - Méthode : multiplication membre par membre puis simplification :

$$\lambda^m p(0) = \mu^m p(m) \iff \rho^m p(0)$$

- **Question 4 : condition d'ergodicité**

$$\rho = \frac{\lambda}{\mu} < 1$$

30

Etude d'une file d'attente

Etude de la M/M/1 : exemple (suite)

- **Question 5 : calcul de p(0) et p(n)**

$\sum p(n) = 1$

$\Leftrightarrow \sum \rho^n p(0) = 1$

$\hookrightarrow p(0) = \frac{1-\rho}{1-\rho^{m+1}}$; $\rho < 1 \Rightarrow p(0) \xrightarrow{n\to\infty} 1-\rho$

- **Question 6 : calcul du débit moyen de la file**

$X = \mu(1-p(0)) = \mu\rho = m\lambda = \frac{m}{20}$

- **Question 7 : calcul de la longueur moyenne de la file**

$Q = \sum n\, p(n) = \frac{\rho}{1-\rho} = \frac{m}{40-m}$

31

Etude d'une file d'attente

Etude de la M/M/1 : exemple (suite)

- **Question 8 : calcul du temps de réponse moyen**
  - Loi de Little : Q = W X

$W = \frac{Q}{X} = \frac{\rho}{1-\rho} \cdot \frac{20}{m}$

$= \frac{20}{40-m}$

32

file M/M/1.

$p(0) = 1-\rho$

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

$X = \mu\rho = m\lambda$ debit

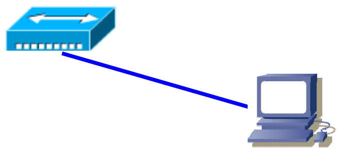
$Q = \frac{\rho}{1-\rho}$ nb clients

$W = \frac{Q}{X}$ tps rép moyen

Etude d'une file d'attente

Exercice 1 :

- On considère un concentrateur sur lequel est connecté un seul terminal.
- La sortie du concentrateur est une ligne de capacité 9600 bits/s
- Le terminal génère en moyenne 1 paquet toutes les 5 s, selon une distribution de Poisson. La longueur des paquets est variable et suit une loi exponentielle de moyenne 600 octets.
- Déterminer le taux d'arrivée des paquets
- Calculer l'intensité du trafic (taux d'utilisation du concentrateur)
- Calculer le nombre moyen de paquets dans la file et le temps moyen de réponse (temps de traversée).
- Déduisez le temps de réponse



33

Etude d'une file d'attente

Exercice 1 (corrigé) :

- Nature du client :

  le paquet (il vaut 600 octets)
- Taux d'arrivée des paquets :

  $\lambda = 1/5 = 0,2$ paquet.s^{-1}
- Taux de service :

  μ = 9600/(600*8) = 2paquets.s^{-1}
- Intensité du trafic :

  $\rho = \frac{\lambda}{\mu} = \frac{0,2}{2} = 0,1$
- Nombre moyen de clients dans le système :

  $Q = \frac{\rho}{1-\rho} = 0,1111$ Paquets
- débit :

  $X = \lambda$ paquets.s^{-1}
- Temps de réponse moyen :

  $W = \frac{Q}{X} = \frac{Q}{\lambda} = 0,555$ s

34

Etude d'une file d'attente

Exercice 1 (suite) :

- On considère désormais 20 terminaux reliés à ce concentrateur. Cinq d'entre eux génèrent en moyenne un paquet d'information toutes les 5 secondes selon une distribution de Poisson, les quinze autres générant un paquet toutes les 25 secondes.
- Longueur des paquets : variable et suit une loi exponentielle de moyenne 600 octets. Le temps de traitement des paquets est considéré négligeable.
- On suppose que le concentrateur peut stocker un nombre arbitraire de paquets.
- Calculer le taux d'arrivée des paquets, le taux d'utilisation de la ligne, le nombre moyen de paquets présents dans le concentrateur, le temps moyen de réponse d'un paquet,.

35

Etude d'une file d'attente

Exercice 1 suite (corrigé) :

- Nature du client :
  paquet
- Taux d'arrivée des 5 terminaux :
  $\lambda_{1-5} = 1/5$
- Taux d'arrivée des 15 terminaux :
  $\lambda_{5-20} = \frac{1}{25} = 0,04$ paquets/s
- Taux d'arrivée globale :
  $\lambda = 5 . \lambda_{1-5} + 15 \lambda_{5-20} = 1,6$ paquets/s
- Taux de service :
  $\mu = \frac{9600}{600 . 8} = 2$ paquets/s
- Intensité du trafic :
  $\rho = \frac{\lambda}{\mu} = \frac{1,6}{2} = 0,8$
- Nombre moyen de clients dans la file :
  $Q = \frac{\rho}{1-\rho} = 4$ paquets
- Temps de réponse :
  $W = \frac{Q}{\lambda} = \frac{4}{\mu\rho} = \frac{4}{1,6} = 2,5$ s

36

Plan

Plan du chapitre 3 :

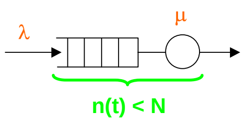
- **Introduction**
- **Présentation**
- **Etude d'une file d'attente**
  - **Présentation**
  - **Approche stochastique**
  - **M/M/1 + exemple + exercice**
  - **M/M/1/N**
  - **M/M/C**
- **Exercices**

37

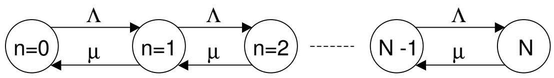
Etude d'une file d'attente

Etude de la M/M/1/N :

- **Identique à une M/M/1 excepté la capacité finie de la file**
  - Processus d'arrivée : processus de Poisson de taux λ
  - Temps de service : variable aléatoire exponentielle de taux μ
  - Capacité de la file : N = nombre MAX de clients (en service + en attente)



- **CMTC associée**



- **Stabilité**
  - CMTC de la file M/M/1/N : irréductible, avec des états récurrents non nuls
    - ➔ Pas de condition de stabilité
    - ➔ Interprétation :
      - ✓ capacité de file limitée, donc même si arrivée des clients bcp plus rapide que le service, les clients "en trop" seront rejetés.
      - ✓ Débit entrant = débit sortant sur un temps très long

38

Etude d'une file d'attente

Etude de la M/M/1/N :

- **Matrice de transition**

$$P = \begin{pmatrix} -\lambda & \lambda & 0 & & & & \\ \mu & -(\mu+\lambda) & \lambda & 0 & & & \\ 0 & \mu & -(\mu+\lambda) & \lambda & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & \mu & -(\mu+\lambda) & \lambda & 0 \\ & & & 0 & \mu & -(\mu+\lambda) & \lambda \\ & & & & 0 & \mu & -\mu \end{pmatrix}$$

- **Régime permanent**
  - Même raisonnement que pour M/M/1, on obtient :
    - ➔ Utilisation les équations d'équilibre
    - ➔ et la condition de normalisation $\sum_{n=0}^{N} p(n) = 1$

$$p(n) = \rho^n p(0) \implies p(0) = \frac{1}{\sum_{n=0}^{N} \rho^n} = \frac{1-\rho}{1-\rho^{N+1}}$$

39

Etude d'une file d'attente

Etude de la M/M/1/N :

- **Paramètres de performances**
  - Débit X :
    - ➔ Soit le taux de départ des clients en sortie du serveur :

$$X_s = P[\text{file non vide}]\,\mu = \sum_{n=1}^{N} p(n)\mu = (1-p(0))\mu = \frac{\rho-\rho^{N+1}}{1-\rho^{N+1}}\mu$$

    - ➔ Soit le taux d'arrivée effectif des clients acceptés dans le système :

$$X_e = P[\text{file non pleine aux ins tan ts d' arrivée}]\,\lambda = \sum_{n=0}^{N-1} p(n)\lambda = (1-p(N))\lambda = \frac{1-\rho^N}{1-\rho^{N+1}}\lambda$$

    - ➔ Puisque $\rho=\lambda/\mu$, $X_e=X_s=X$

  - Longueur moyenne de la file d'attente (= nombre moyen de clients)

$$Q = \sum n\, p(n) = \cdots$$

  - Temps de réponse moyen (ou temps de séjour moyen) pour les clients admis dans le système :

$$W = \frac{Q}{X} = \cdots$$

40

Plan

Plan du chapitre 3 :

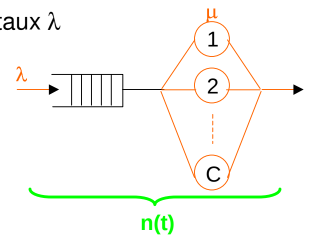
- **Introduction**
- **Présentation**
- **Etude d'une file d'attente**
  - **Présentation**
  - **Approche stochastique**
  - **M/M/1 + exemple + exercice**
  - **M/M/1/N**
  - **M/M/C**
- **Exercices**

41

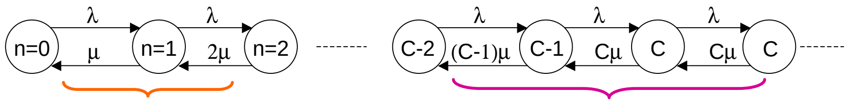
Etude d'une file d'attente

Etude de la M/M/C :

- **Identique à une M/M/1 excepté le nombre C de serveurs identiques et indépendants les uns des autres**
  - Processus d'arrivée : processus de Poisson de taux λ
  - Temps de service : exponentiel de taux μ (pour chacun des serveurs)
  - Capacité de la file : infinie



- **CMTC associée**



chaque serveur s'occupe d'1 client

Même si le nombre de clients augmente, on reste à C serveurs

$$\begin{cases} \lambda(0) = \lambda(1) = \cdots = \lambda \\ \mu(1) = \mu; \mu(2) = 2\mu; \cdots; \mu(C) = C\mu = \mu(C+1) = \mu(C+2) = \cdots \end{cases}$$

42

Etude d'une file d'attente

Etude de la M/M/C :

- **Stabilité (condition d'ergodicité)**
  - Nombre moyen de client arrivant dans la file par unité de temps < nombre moyen de clients pouvant être traités par les C serveurs par unité de temps
  - D'où : $\lambda < C\mu$
- **Régime permanent :** écriture des équations d'équilibre (de frontière)

$$\begin{cases} p(n-1)\lambda = p(n)n\mu & \text{pour } n = 1,\cdots,C-1 \\ p(n-1)\lambda = p(n)C\mu & \text{pour } n = C, C+1,\cdots \end{cases} \qquad \text{or } \rho = \frac{\lambda}{\mu}$$

$$\begin{cases} p(n) = \frac{\rho}{n}p(n-1) & \text{pour } n = 1,\cdots,C-1 \\ p(n) = \frac{\rho}{C}p(n-1) & \text{pour } n = C, C+1,\cdots \end{cases}$$

en fonction de p(0)

$$\begin{cases} p(n) = \frac{\rho}{n!}p(0) & \text{pour } n = 1,\cdots,C-1 \\ p(n) = \frac{\rho^n}{C!\,C^{n-C}}p(0) & \text{pour } n = C, C+1,\cdots \end{cases} \quad \textbf{et} \quad p(0) = \frac{1}{\sum_{n=0}^{C-1}\frac{\rho^n}{n!} + \frac{\rho^C}{(C-1)!\,(C-\rho)}}$$

Condition de normalisation

Etude d'une file d'attente

Etude de la M/M/C :

- **Paramètres de performances**
  - Débit : X

$$X = \underbrace{\sum_{n=1}^{C-1} p(n)n\mu}_{\text{Service de taux } n\mu,\ \text{pour } n < C} + \underbrace{\sum_{n=C}^{\infty} p(n)C\mu}_{\text{Service de taux } C\mu,\ \text{pour } n > C} \Longrightarrow \boxed{X = \lambda}$$

  - Temps moyen de séjour W
    - = temps moyen d'attente dans le buffer de la file + temps moyen de service
    - $$W = \frac{\rho^C}{\mu(C-1)!\,(C-\rho)^2}p(0) + \frac{1}{\mu}$$
  - Nombre moyen de client Q
    - Application de la loi de Little à l'ensemble de la file
      - ✓ $Q = W\,X$

Etude d'une file d'attente

Processus de naissance et de mort :

- **Extension de la M/M/1 - file avec dépendance de l'état**
  - Distribution du temps de service dépend du nombre de clients dans la file :
    - ➔ n clients :
      - ✓ Taux d'arrivée : $\lambda(n)$ (poissonien)
      - ✓ Taux de service : $\mu(n)$ (exponentiel)
- **Définition**
  - Processus de naissance et de mort = processus de Markov dans lequel 2 transitions seulement sont possibles
    - ➔ 1 mort : n → n-1
    - ➔ 1 naissance : n → n+1
- **Matrice de transition**

$$P = \begin{pmatrix} -\lambda(0) & \lambda(0) & 0 & 0 & \cdots \\ \mu(1) & -\lambda(1)-\mu(1) & \lambda(1) & 0 & \cdots \\ 0 & \mu(2) & -\lambda(2)-\mu(2) & \lambda(2) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

45

Etude d'une file d'attente

Processus de naissance et de mort :

- **Equation d'équilibre (de balance)**

$$p(n) = \frac{\lambda(0)\lambda(1)\cdots\lambda(n-1)}{\mu(1)\mu(2)\cdots\mu(n)}p(0)$$

- **Condition d'existence d'une solution stationnaire**

$$p(0) \neq 0 \quad \textbf{et} \quad \sum_{n=0}^{\infty} p(n) = 1$$

$$p(0) = 1 + \sum_{n=0}^{\infty} \frac{\lambda(0)\cdots\lambda(n)}{\mu(1)\cdots\mu(n+1)}$$

46

Plan

Plan du chapitre 3 :

- Introduction
- Présentation
- Etude d'une file d'attente
- **Exercices**

47

Etude d'une file d'attente

Exercice 2 :

- Considérons le central téléphonique d'une entreprise comprenant N lignes d'entrées
- Arrivées des appels : selon un processus de Poisson de taux λ=1
- Durée des communications distribuée exponentiellement de taux μ. Elle est en moyenne de 2 unités de temps (μ=0.5).
- Les appels trouvant N lignes occupées sont perdus.
- Question : calculer le nombre de lignes nécessaires pour que le nombre d'appels trouvant les N lignes occupées soit inférieur à 10%.
- Caractéristique de la file

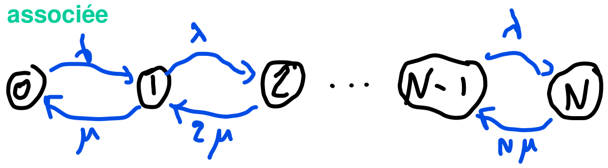
  Processus d'arrivée : exponentiel, de taux lambda
  Processus de service : exponentiel, de taux mu
  Nb de serveur : N
  Capacité de la file : N
  Discipline de services : FIFO

48

Etude d’une file d’attente

Exercice 2 (suite) :

- **CMTC associée**



- **Matrice de transition**

$$
\begin{array}{c|cccccc}
 & 0 & 1 & 2 & \cdots & N-1 & N \\
\hline
0 & -\lambda & \lambda & 0 & \cdots & 0 & 0 \\
1 & \mu & -(\mu+\lambda) & \lambda & \cdots & 0 & 0 \\
2 & 0 & 2\mu & -(2\mu+\lambda) & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
N-1 & 0 & 0 & 0 & \cdots & -((N-1)\mu+\lambda) & \lambda \\
N & 0 & 0 & 0 & \cdots & \mu & -N\mu
\end{array}
$$

Etude d’une file d’attente

Exercice 2 (suite) :

- **Equation d'équilibre**

$$-\lambda p(0) + \mu p(1) = 0$$
$$\lambda p(0) - (\mu+\lambda) p(1) + 2\mu p(2) = 0$$
$$\vdots$$

$$\Leftrightarrow \quad \lambda p(0) = \mu p(1)$$
$$(\mu+\lambda) p(1) = \lambda p(0) + 2\mu p(2)$$
$$\vdots$$

$$\Leftrightarrow \quad \lambda p(0) = \mu p(1)$$
$$\lambda p(1) = 2\mu p(2)$$
$$\lambda p(2) = 3\mu p(3)$$
$$\vdots$$

On multiplie membre à membre :

$$\lambda^m . p(0) . p(1) . p(2) \ldots p(m-1) = m! \, \mu . p(1) . p(2) \ldots p(N)$$

$$\lambda^m p(0) = m! \, \mu^m p(m)$$

$$p(m) = \frac{1}{m!} \cdot \left(\frac{\lambda}{\mu}\right)^m \cdot p(0)$$

Etude d’une file d’attente

Exercice 2 (suite) :

- **Expression de p(0)**

$\sum_{n=0}^{N} p(n) = 1$ or, $p(n) = \frac{e^n}{n!} p(0)$

par identif. : $p(0) = \left(\sum \frac{e^n}{n!} \right)^{-1}$

- **Nombre de lignes nécessaires**
  - Si les N lignes sont occupées, il y a N clients dans la file : p(n)
  - Probabilité de perdre un client = probabilité d'avoir N clients (file pleine)
  - Calcul des p(n) pour différentes valeurs de n et si p(n)<10% alors n=N

51

Etude d’une file d’attente

Exercice 2 (suite) :

- **Nombre de lignes nécessaires**

$p(1) = \frac{1}{1!} \cdot 2^1 \left(\frac{e^0}{0!} + \frac{e^1}{1!} \right)^{-1}$

$= \frac{2}{3} \rightarrow 66,7\%$

$p(2) = 40\%$

$p(3) = 21\%$

$p(4) = 0,95\% < 10 \Rightarrow$ 4 lignes nécessaires

52

Etude d’une file d’attente

Exercice 1 (suite) :

- On considère toujours 20 terminaux reliés à un concentrateur (ligne de capacité 9600 bits/s). Cinq d’entre eux génèrent en moyenne un paquet d’information toutes les 5 secondes selon une distribution de Poisson, les quinze autres générant un paquet toutes les 25 secondes.
- Longueur des paquets : variable et suit une loi exponentielle de moyenne 600 octets. Le temps de traitement des paquets est considéré négligeable.
- On souhaite diminuer le temps moyen de réponse des paquets, soit en doublant la capacité de transmission de la ligne, soit en mettant en parallèle une seconde ligne de transmission de même débit.
- Evaluer ces deux possibilités.

53

Etude d’une file d’attente

Exercice 1 (suite) - corrections : si capacité doublée

- Nature de la file :
- Taux d'arrivée globale :
- Taux de service :
- Intensité du trafic :
- Nombre moyen de clients dans la file :
- Temps de réponse :

54

Etude d’une file d’attente

Exercice 1 (suite) - corrections : si doublage de la ligne

- Nature de la file :
- Taux d'arrivée globale :
- Taux de service :
- Intensité du trafic :
- Taux d'utilisation de chaque serveur :
- Temps de réponse :

avec

55

Etude d’une file d’attente

Exercice 1 (suite) - corrections : comparaison

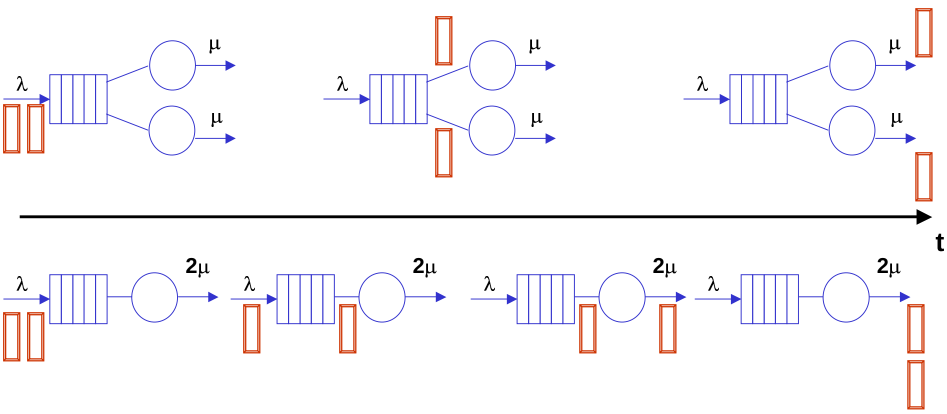
- Temps de réponse si capacité de la ligne doublée :
- Temps de réponse si doublage de la ligne :
- Conclusion :

56

Etude d’une file d’attente

Exercice 1 (suite) - corrections : comparaison - explications

- **Arrivée de messages très rapprochée dans le temps :**

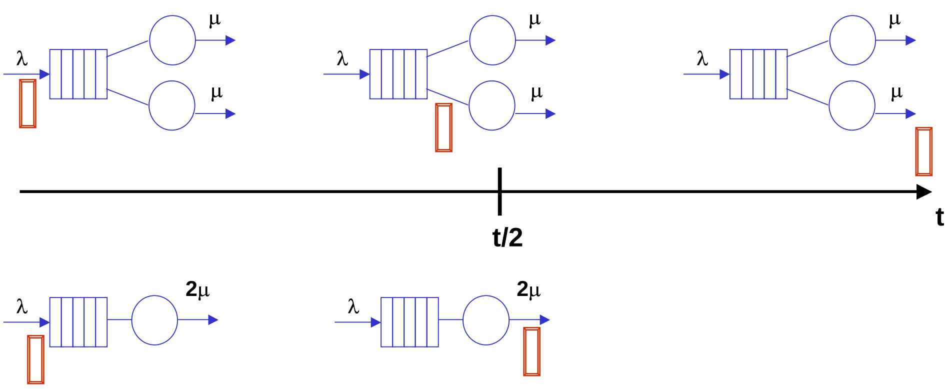


Performances temporelles équivalentes

Etude d’une file d’attente

Exercice 1 (suite) - corrections : comparaison - explications

- **Arrivée de messages très espacée dans le temps :**



Performance "doublée" avec une ligne à capacité doublée